PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
C12N 15/29, C07K 14/415, C12N 15/82, 15/11, 1/21, G01N 33/53, A01H 5/00, C12N 7/01, C12Q 1/68, A01N 65/00

A2

(11) International Publication Number: WO 00/60087

(43) International Publication Date: 12 October 2000 (12.10.00)

(21) International Application Number: PCT/US00/09106

(22) International Filing Date: 6 April 2000 (06.04.00)

(30) Priority Data:
60/128,192 7 April 1999 (07.04.99) US

(71) Applicant *(for all designated States except US)*: E. I. DU PONT DE NEMOURS AND COMPANY [US/US]; 1007 Market Street, Wilmington, DE 19898 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: KLEIN, Theodore, M. [US/US]; 2229 Rosewood Drive, Wilmington, DE 19810 (US). WENG, Zude [CN/US]; Apartment 1B, 9122 Lincoln Drive, Des Plaines, IL 60016 (US). CAHOON, Rebecca, E. [US/US]; 2331 West 18th Street, Wilmington, DE 19806 (US).

(74) Agent: GEIGER, Kathleen, W.; E.I. du Pont de Nemours and Company, Legal Patent Records Center, 1007 Market Street, Wilmington, DE 19898 (US).

(81) Designated States: AE, AL, AU, BA, BB, BG, BR, CA, CN, CR, CU, CZ, DM, EE, GD, GE, HR, HU, ID, IL, IN, IS, JP, KP, KR, LC, LK, LR, LT, LV, MG, MK, MN, MX, NO, NZ, PL, RO, SG, SI, SK, SL, TR, TT, UA, US, UZ, VN, YU, ZA, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: CELL CYCLE GENES IN PLANTS

```
Chenopodium       M-----------------------------------------AAA----------
Arabidopsis       MGKYIRKSKIDGAGAGAGGGGGGGGGESSIALMDVVSPSSSSSLGVLTRAKSLALQQQQ
Nematode          MNGTESECERQPAQDGQKMENEVCEGVANTS--------EGGSDGKISETENETGDLEEIP
SEQ ID NO:10      ----------------------------------------------------------
SEQ ID NO:12      ----------------------------------------------------------
SEQ ID NO:14      IGLVSMVTEQPSED---------DGGVNFQ--------DGGGSSAALLDKNAGEEIELVA
SEQ ID NO:16      MGKYMRKAKASSE----------------VVIMDVAAAP-----LGVRTRARALALQRLQ
SEQ ID NO:18      ----------------------------------------------------------
SEQ ID NO:20      ARVTPARV----------------------------------------------------
SEQ ID NO:22      ----------------------------------------------------------
SEQ ID NO:24      T-------------------------------------------FG-VG----------
SEQ ID NO:26      M-----------------------------------------SAQVGVRTRAQA-------
SEQ ID NO:28      ----------------------------------------------------------
SEQ ID NOs:30/32  MG---KYMKKSKI---------------AGDVAAVIME--AFPPKSHLGVRTRAKTLALQ
SEQ ID NO:34      MGKYMRKPKVSGE----------------VAVHEVAAAP----LGVRTRARALAMQR--

Chenopodium       ---------------ATPTSSPAKKIKKVSKSSTNIP------QLRSRKKKLSAPENF--
Arabidopsis       QRCLLQKPSSPSSLPPTSASPWPPSKQKMKKKQQQMNDCGSYLQLRSRRLQKKPPIVVIR
Nematode          GAEVVADPH-----------ERASNEAEIQNDSEILDVGETNTSGSENEENTSQNKMED
SEQ ID NO:10      ----------------------------------------------------A---C--
SEQ ID NO:12      ----------------------------------------------------GSRTC--
SEQ ID NO:14      GVRALQVP-------------LTGNGSDASRESEMRKFMEVLTVREKNSEVDTQEVSKE
SEQ ID NO:16      EQ-----------------------------QTQMEEGAGGEYLELRNRRLEKLPPPPATT
SEQ ID NO:18      ----------------------------------------------------GK-----
SEQ ID NO:20      ----------------------------------------Y-------------------
SEQ ID NO:22      ----------------------------------------------------------
SEQ ID NO:24      ----------------------------------------------------------
SEQ ID NO:26      ----------ALAMZAVSSAEPSSKRKKISRSTNQEP------KL--SKTPRTSSSSA--
SEQ ID NO:28      ----------------STEEHITKTKS----------------------------------
SEQ ID NOs:30/32  MR----------------TTSPDPSA-----------------YLQLRSRRLLKLPPTP
SEQ ID NO:34      QP----------------------------QGAAVAKDQGEYLELASRKLEKLPPPPPAA
```

(57) Abstract

This invention relates to an isolated nucleic acid fragment encoding a cyclin dependent kinase inhibitor (CDKI). The invention also relates to the construction of a chimeric gene encoding all or a substantial portion of the CDKI, in sense or antisense orientation, wherein expression of the chimeric gene results in production of altered levels of the CDKI in a transformed host cell.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

TITLE

CELL CYCLE GENES IN PLANTS

This application claims the benefit of U.S. Provisional Application No. 60/128,192, filed April 7, 1999.

FIELD OF THE INVENTION

This invention is in the field of plant molecular biology. More specifically, this invention pertains to nucleic acid fragments encoding a cyclin dependent kinase inhibitor in plants and seeds.

BACKGROUND OF THE INVENTION

Cells divide in a carefully orchestrated series of events that starts with the decision to duplicate the nuclear DNA complement and ends with the physical separation of the two daughter cells. The cycle of events is normally divided into four stages, G1, S, G2 and M. The S and M phases (DNA synthesis and cell mitosis) are largely constant in their length and progression, while the G1 and G2 phases are highly variable. The decision to enter S (and the requisite M phase that will follow) sets the cycle into an irreversible forward direction. If the cell is not capable of completing the cycle, due to insufficient cellular resources, the result is usually cell death. Therefore, the decision making process to initiate a cell division must be carefully regulated to monitor many diverse cellular properties and resources.

Cell cycle timing is largely controlled by the concentration of proteins called cyclins. The cyclins are a class of proteins, some of which serve as the cellular clock that regulates the forward progression of the cell cycle. Mitotic cyclins associate with a class of kinases (cyclin-dependent kinases, CDKs) allowing for the specific activation of the kinase activity, which in turn causes the phosphorylation and activation of several key transcriptional and translational activators. These activators turn on the complement of genes that are required to complete DNA synthesis and cytokinesis. Regulation of this process occurs at multiple levels. The number and complexity of the regulatory controls working on the cell cycle precludes an exhaustive discussion here. Indeed, many of the forms of regulation may be specific for a species, developmental process, or cell-type. The general mechanisms for the activation of the CDKs requires both association with a cyclin, as well as phosphorylation and dephosphorylation events. There are also phosphatase activities that work to promote and inhibit the cell cycle. Also, several classes of transcriptional inhibitors (often termed tumor suppressors) serve to prevent inappropriate activation of gene expression.

One central, and universal, form of regulation of the cell cycle is the inhibition of the CDKs that regulate the timing of cell cycle events. One class of proteins are called CDK inhibitors (CDKIs or CKIs). These proteins are generally not kinases or phosphatases, but instead associate with either the cyclin or kinase subunits and prevent

the action of activating kinases or phosphatases. Manipulation of these inhibitors can lead to enhancement of cell division and growth, or blockage of cell division and death. Understanding the contribution of individual members of this family will allow for a clearer picture of cell cycle regulation in plants. Also, plant herbicide, and/or plant growth promoting compounds may be discovered that use CDKIs as their targets. Nucleic acid sequences are described herein that encode CDKIs from corn, soybean, rice, and wheat.

Several plant CDKI cDNA sequences have been reported, seven from *Arabidopsis*, one from *Chenopodium*, and one from alfalfa (as described in PCT Publication Nos. WO 99/14331 and WO 99/64599, published on March 25, 1999 and December 16, 1999, respectively, the disclosures of which are hereby incorporated by reference).

SUMMARY OF THE INVENTION

The present invention concerns an isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of: (a) a first nucleotide sequence encoding a polypeptide of at least 50 amino acids having at least 75% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, and 34, or (b) a second nucleotide sequence comprising the complement of the first nucleotide sequence.

In a second embodiment, it is preferred that the isolated polynucleotide of the claimed invention comprises a first nucleotide sequence which comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33 that codes for the polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, and 34.

In a third embodiment, this invention concerns an isolated polynucleotide comprising a nucleotide sequence of at least one of 60 (preferably at least one of 40, most preferably at least one of 30) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33 and the complement of such nucleotide sequences.

In a fourth embodiment, this invention relates to a chimeric gene comprising an isolated polynucleotide of the present invention operably linked to at least one suitable regulatory sequence.

In a fifth embodiment, the present invention concerns a host cell comprising a chimeric gene of the present invention or an isolated polynucleotide of the present invention. The host cell may be eukaryotic, such as a yeast or a plant cell, or prokaryotic, such as a bacterial cell. The present invention also relates to a virus, preferably a baculovirus, comprising an isolated polynucleotide of the present invention or a chimeric gene of the present invention.

In a sixth embodiment, the invention also relates to a process for producing a host cell comprising a chimeric gene of the present invention or an isolated polynucleotide of the

present invention, the process comprising either transforming or transfecting a compatible host cell with a chimeric gene or isolated polynucleotide of the present invention.

In a seventh embodiment, the invention concerns a CDKI polypeptide of at least 50 amino acids comprising at least 75% identity based on the Clustal method of alignment compared to a polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, and 34.

In an eighth embodiment, the invention relates to a method of selecting an isolated polynucleotide that affects the level of expression of a CDKI polypeptide or enzyme activity in a host cell, preferably a plant cell, the method comprising the steps of: (a) constructing an isolated polynucleotide of the present invention or a chimeric gene of the present invention; (b) introducing the isolated polynucleotide or the chimeric gene into a host cell; (c) measuring the level of the CDKI polypeptide or enzyme activity in the host cell containing the isolated polynucleotide; and (d) comparing the level of the CDKI polypeptide or enzyme activity in the host cell containing the isolated polynucleotide with the level of the CDKI polypeptide or enzyme activity in the host cell that does not contain the isolated polynucleotide.

In a ninth embodiment, the invention concerns a method of obtaining a nucleic acid fragment encoding a substantial portion of a CDKI polypeptide, preferably a plant CDKI polypeptide, comprising the steps of: synthesizing an oligonucleotide primer comprising a nucleotide sequence of at least one of 60 (preferably at least one of 40, most preferably at least one of 30) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33, and the complement of such nucleotide sequences; and amplifying a nucleic acid fragment (preferably a cDNA inserted in a cloning vector) using the oligonucleotide primer. The amplified nucleic acid fragment preferably will encode a substantial portion of a CDKI amino acid sequence.

In a tenth embodiment, this invention relates to a method of obtaining a nucleic acid fragment encoding all or a substantial portion of the amino acid sequence encoding a CDKI polypeptide comprising the steps of: probing a cDNA or genomic library with an isolated polynucleotide of the present invention; identifying a DNA clone that hybridizes with an isolated polynucleotide of the present invention; isolating the identified DNA clone; and sequencing the cDNA or genomic fragment that comprises the isolated DNA clone.

In an eleventh embodiment, this invention concerns a composition, such as a hybridization mixture, comprising an isolated polynucleotide of the present invention.

In a twelfth embodiment, this invention concerns a method for positive selection of a transformed cell comprising: (a) transforming a host cell with the chimeric gene of the present invention or a construct of the present invention; and (b) growing the transformed host cell, preferably a plant cell, such as a monocot or a dicot, under conditions which allow

expression of the CDKI polynucleotide in an amount sufficient to complement a null mutant to provide a positive selection means.

A further embodiment of the instant invention is a method for evaluating at least one compound for its ability to inhibit the activity of a cyclin dependent kinase inhibitor, the method comprising the steps of: (a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding a cyclin dependent kinase inhibitor polypeptide, operably linked to at least one suitable regulatory sequence; (b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results in production of the cell-cycle regulatory gene in the transformed host cell; (c) optionally purifying the cyclin dependent kinase inhibitor polypeptide expressed by the transformed host cell; (d) treating the cyclin dependent kinase inhibitor polypeptide with a compound to be tested; and (e) comparing the activity of the cyclin dependent kinase inhibitor polypeptide that has been treated with a test compound to the activity of an untreated cyclin dependent kinase inhibitor polypeptide, thereby selecting compounds with potential for inhibitory activity.

BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE LISTINGS

The invention can be more fully understood from the following detailed description, the accompanying drawings and Sequence Listing which form a part of this application.

Figure 1 shows a comparison of the amino acid sequences of the plant CDKI amino acid sequences. The amino acid sequences from SEQ ID NOs:10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, and 34 are shown compared to the amino acid sequences of the closest art, the short-day flowering plant *Chenopodium rubrum* [NCBI Accession No. gi 265281, SEQ ID NO:35], the nematode *Caenorhabditis elegans* [NCBI Accession No. gi 2731583, SEQ ID NO:36], and the flowering weed *Arabidopsis thaliana* [NCBI Accession No. gi 2914702, SEQ ID NO:37]. The *Arabidopsis* gene is an unidentified open reading frame from a genomic sequencing project (Lin X., et al. (1999) *Nature 402*:761-768), this ORF appears to contain an unidentified intron at position 30689-30802 of the genomic clone. Removal of this sequence, which contains the conserved GT/AG intron border sequences, removes 38 amino acids and brings together the sequences that allow for an identification of this gene as a CDKI. The corrected nucleotide sequence for the mRNA encoding this gene is presented in SEQ ID NO:38 and the translation is shown in SEQ ID NO:39. SEQ ID NO:39 is used in the alignment shown in Figure 1.

Table 1 lists the polypeptides that are described herein, the designation of the cDNA clones that comprise the nucleic acid fragments encoding polypeptides representing all or a substantial portion of these polypeptides, and the corresponding identifier (SEQ ID NO:) as used in the attached Sequence Listing. The sequence descriptions and Sequence Listing

attached hereto comply with the rules governing nucleotide and/or amino acid sequence disclosures in patent applications as set forth in 37 C.F.R. §1.821-1.825.

TABLE 1

Plant Cyclin dependent kinase inhibitor Genes

CDKI Gene [Plant Source] In Provisional (yes/no/partial)	Clone Designation	SEQ ID NO: (Nucleotide)	(Amino Acid)
Corn [*Zea mays*] (yes)	csi1n.pk0050.e6	1	2
Rice [*Oryza sativa*] (yes)	rsr9n.pk003.g12	3	4
Soybean [*Glycine max*] (yes)	sl2.pk0008.d2	5	6
Wheat [*Triticum aestivum*] (yes)	wre1n.pk0031.g2	7	8
Corn [*Zea mays*] (partial)	csi1n.pk0050.e6:fis	9	10
Corn [*Zea mays*] (no)	contig of: cr1n.pk0052.f11, cr1n.pk0195.b3, p0095.cwsbc53r	11	12
Corn [*Zea mays*] (no)	p0023.clrag76r	13	14
Corn [*Zea mays*] (no)	contig of: cen3n.pk0013.c9, cen3n.pk0115.a6, cen3n.pk0151.d2, cs1.pk0068.c12, cta1n.pk0070.d4, p0058.chpbm23rb	15	16
Corn [*Zea mays*] (no)	contig of: cbn2n.pk0010.h7, p0086.cbsaj18f, p0086.cbsaj18r, p0086.cbsaj18rb	17	18
Rice [*Oryza sativa*] (partial)	rsr9n.pk003.g12:fis	19	20
Rice [*Oryza sativa*] (no)	rds2c.pk008.o24	21	22
Soybean [*Glycine max*] (partial)	sl2.pk0008.d2:fis	23	24
Soybean [*Glycine max*] (no)	sdp4c.pk025.k23	25	26
Soybean [*Glycine max*] (no)	sl2.pk0117.h4	27	28
Soybean [*Glycine max*] (no)	sls1c.pk007.h20-5'	29	30
Soybean [*Glycine max*] (no)	sls1c.pk007.h20-3'	31	32
Wheat [*Triticum aestivum*] (part.)	wre1n.pk0031.g2:fis	33	34

The soybean clone sls1c.pk007.h20 represents a full length CDKI sequence. Sequences from the 5'-end of the clone, and from the 3'-end of the clone, are presented in SEQ ID NOs:29 and 31, respectively.

The Sequence Listing contains the one letter code for nucleotide sequence characters and the three letter codes for amino acids as defined in conformity with the IUPAC-IUBMB

standards described in *Nucleic Acids Res. 13*:3021-3030 (1985) and in the *Biochemical J. 219 (No. 2)*:345-373 (1984) which are herein incorporated by reference. The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

DETAILED DESCRIPTION OF THE INVENTION

In the context of this disclosure, a number of terms shall be utilized. The terms "polynucleotide", "polynucleotide sequence", "nucleic acid sequence", and "nucleic acid fragment"/"isolated nucleic acid fragment" are used interchangeably herein. These terms encompass nucleotide sequences and the like. A polynucleotide may be a polymer of RNA or DNA that is single- or double-stranded, that optionally contains synthetic, non-natural or altered nucleotide bases. A polynucleotide in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA, synthetic DNA, or mixtures thereof. An isolated polynucleotide of the present invention may include at least one of 60 contiguous nucleotides, preferably at least one of 40 contiguous nucleotides, most preferably one of at least 30 contiguous nucleotides derived from SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33, or the complement of such sequences.

The term "isolated polynucleotide" refers to a polynucleotide that is substantially free from other nucleic acid sequences, such as and not limited to other chromosomal and extrachromosomal DNA and RNA. Isolated polynucleotides may be purified from a host cell in which they naturally occur. Conventional nucleic acid purification methods known to skilled artisans may be used to obtain isolated polynucleotides. The term also embraces recombinant polynucleotides and chemically synthesized polynucleotides.

The term "recombinant" means, for example, that a nucleic acid sequence is made by an artificial combination of two otherwise separated segments of sequence, e.g., by chemical synthesis or by the manipulation of isolated nucleic acids by genetic engineering techniques.

As used herein, "contig" refers to a nucleotide sequence that is assembled from two or more constituent nucleotide sequences that share common or overlapping regions of sequence homology. For example, the nucleotide sequences of two or more nucleic acid fragments can be compared and aligned in order to identify common or overlapping sequences. Where common or overlapping sequences exist between two or more nucleic acid fragments, the sequences (and thus their corresponding nucleic acid fragments) can be assembled into a single contiguous nucleotide sequence.

As used herein, "substantially similar" refers to nucleic acid fragments wherein changes in one or more nucleotide bases results in substitution of one or more amino acids, but do not affect the functional properties of the polypeptide encoded by the nucleotide sequence. "Substantially similar" also refers to nucleic acid fragments wherein changes in one or more nucleotide bases does not affect the ability of the nucleic acid fragment to mediate alteration of gene expression by gene silencing through for example antisense or co-

suppression technology. "Substantially similar" also refers to modifications of the nucleic acid fragments of the instant invention such as deletion or insertion of one or more nucleotides that do not substantially affect the functional properties of the resulting transcript vis-à-vis the ability to mediate gene silencing or alteration of the functional properties of the resulting protein molecule. It is therefore understood that the invention encompasses more than the specific exemplary nucleotide or amino acid sequences and includes functional equivalents thereof. The terms "substantially similar" and "corresponding substantially" are used interchangeably herein.

Substantially similar nucleic acid fragments may be selected by screening nucleic acid fragments representing subfragments or modifications of the nucleic acid fragments of the instant invention, wherein one or more nucleotides are substituted, deleted and/or inserted, for their ability to affect the level of the polypeptide encoded by the unmodified nucleic acid fragment in a plant or plant cell. For example, a substantially similar nucleic acid fragment representing at least one of 30 contiguous nucleotides derived from the instant nucleic acid fragment can be constructed and introduced into a plant or plant cell. The level of the polypeptide encoded by the unmodified nucleic acid fragment present in a plant or plant cell exposed to the substantially similar nucleic fragment can then be compared to the level of the polypeptide in a plant or plant cell that is not exposed to the substantially similar nucleic acid fragment.

For example, it is well known in the art that antisense suppression and co-suppression of gene expression may be accomplished using nucleic acid fragments representing less than the entire coding region of a gene, and by using nucleic acid fragments that do not share 100% sequence identity with the gene to be suppressed. Moreover, alterations in a nucleic acid fragment which result in the production of a chemically equivalent amino acid at a given site, but do not effect the functional properties of the encoded polypeptide, are well known in the art. Thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also be expected to produce a functionally equivalent product. Nucleotide changes which result in alteration of the N-terminal and C-terminal portions of the polypeptide molecule would also not be expected to alter the activity of the polypeptide. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products. Consequently, an isolated polynucleotide comprising a nucleotide sequence of at least one of 60 (preferably at least one of 40, most preferably at least one of 30) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23,

25, 27, 29, 31, and 33, and the complement of such nucleotide sequences may be used in methods of selecting an isolated polynucleotide that affects the expression of a CDKI polypeptide in a host cell. A method of selecting an isolated polynucleotide that affects the level of expression of a polypeptide in a virus or in a host cell (eukaryotic, such as plant or yeast, prokaryotic such as bacterial) may comprise the steps of: constructing an isolated polynucleotide of the present invention or a chimeric gene of the present invention; introducing the isolated polynucleotide or the chimeric gene into a host cell; measuring the level of a polypeptide or enzyme activity in the host cell containing the isolated polynucleotide; and comparing the level of a polypeptide or enzyme activity in the host cell containing the isolated polynucleotide with the level of a polypeptide or enzyme activity in a host cell that does not contain the isolated polynucleotide.

Moreover, substantially similar nucleic acid fragments may also be characterized by their ability to hybridize. Estimates of such homology are provided by either DNA-DNA or DNA-RNA hybridization under conditions of stringency as is well understood by those skilled in the art (Hames and Higgins, Eds. (1985) Nucleic Acid Hybridisation, IRL Press, Oxford, U.K.). Stringency conditions can be adjusted to screen for moderately similar fragments, such as homologous sequences from distantly related organisms, to highly similar fragments, such as genes that duplicate functional enzymes from closely related organisms. Post-hybridization washes determine stringency conditions. One set of preferred conditions uses a series of washes starting with 6X SSC, 0.5% SDS at room temperature for 15 min, then repeated with 2X SSC, 0.5% SDS at 45°C for 30 min, and then repeated twice with 0.2X SSC, 0.5% SDS at 50°C for 30 min. A more preferred set of stringent conditions uses higher temperatures in which the washes are identical to those above except for the temperature of the final two 30 min washes in 0.2X SSC, 0.5% SDS was increased to 60°C. Another preferred set of highly stringent conditions uses two final washes in 0.1X SSC, 0.1% SDS at 65°C.

Substantially similar nucleic acid fragments of the instant invention may also be characterized by the percent identity of the amino acid sequences that they encode to the amino acid sequences disclosed herein, as determined by algorithms commonly employed by those skilled in this art. Suitable nucleic acid fragments (isolated polynucleotides of the present invention) encode polypeptides that are at least about 70% identical, preferably at least about 80% identical to the amino acid sequences reported herein. Preferred nucleic acid fragments encode amino acid sequences that are about 85% identical to the amino acid sequences reported herein. More preferred nucleic acid fragments encode amino acid sequences that are at least about 90% identical to the amino acid sequences reported herein. Most preferred are nucleic acid fragments that encode amino acid sequences that are at least about 95% identical to the amino acid sequences reported herein. Suitable nucleic acid fragments not only have the above identities but typically encode a polypeptide having at

least 50 amino acids, preferably at least 100 amino acids, more preferably at least 150 amino acids, still more preferably at least 200 amino acids, and most preferably at least 250 amino acids. Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS.* 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

A "substantial portion" of an amino acid or nucleotide sequence comprises an amino acid or a nucleotide sequence that is sufficient to afford putative identification of the protein or gene that the amino acid or nucleotide sequence comprises. Amino acid and nucleotide sequences can be evaluated either manually by one skilled in the art, or by using computer-based sequence comparison and identification tools that employ algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul et al. (1993) *J. Mol. Biol. 215*:403-410; see also www.ncbi.nlm.nih.gov/BLAST/). In general, a sequence of ten or more contiguous amino acids or thirty or more contiguous nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene-specific oligonucleotide probes comprising 30 or more contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., *in situ* hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12 or more nucleotides may be used as amplification primers in PCR in order to obtain a particular nucleic acid fragment comprising the primers. Accordingly, a "substantial portion" of a nucleotide sequence comprises a nucleotide sequence that will afford specific identification and/or isolation of a nucleic acid fragment comprising the sequence. The instant specification teaches amino acid and nucleotide sequences encoding polypeptides that comprise one or more particular plant proteins. The skilled artisan, having the benefit of the sequences as reported herein, may now use all or a substantial portion of the disclosed sequences for purposes known to those skilled in this art. Accordingly, the instant invention comprises the complete sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above.

"Codon degeneracy" refers to divergence in the genetic code permitting variation of the nucleotide sequence without effecting the amino acid sequence of an encoded polypeptide. Accordingly, the instant invention relates to any nucleic acid fragment comprising a nucleotide sequence that encodes all or a substantial portion of the amino acid sequences set forth herein. The skilled artisan is well aware of the "codon-bias" exhibited by a specific host cell in usage of nucleotide codons to specify a given amino acid.

Therefore, when synthesizing a nucleic acid fragment for improved expression in a host cell, it is desirable to design the nucleic acid fragment such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

"Synthetic nucleic acid fragments" can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form larger nucleic acid fragments which may then be enzymatically assembled to construct the entire desired nucleic acid fragment. "Chemically synthesized", as related to a nucleic acid fragment, means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of nucleic acid fragments may be accomplished using well established procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. Accordingly, the nucleic acid fragments can be tailored for optimal gene expression based on optimization of the nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon usage is biased towards those codons favored by the host. Determination of preferred codons can be based on a survey of genes derived from the host cell where sequence information is available.

"Gene" refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. "Native gene" refers to a gene as found in nature with its own regulatory sequences. "Chimeric gene" refers any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. "Endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign gene" refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A "transgene" is a gene that has been introduced into the genome by a transformation procedure.

"Coding sequence" refers to a nucleotide sequence that codes for a specific amino acid sequence. "Regulatory sequences" refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, and polyadenylation recognition sequences.

"Promoter" refers to a nucleotide sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. The promoter sequence consists of proximal and more distal upstream

elements, the latter elements often referred to as enhancers. Accordingly, an "enhancer" is a nucleotide sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. Promoters may be derived in their entirety from a native gene, or may be composed of different elements derived from different promoters found in nature, or may even comprise synthetic nucleotide segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. Promoters which cause a nucleic acid fragment to be expressed in most cell types at most times are commonly referred to as "constitutive promoters". New promoters of various types useful in plant cells are constantly being discovered; numerous examples may be found in the compilation by Okamuro and Goldberg (1989) *Biochemistry of Plants 15*:1-82. It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, nucleic acid fragments of different lengths may have identical promoter activity.

"Translation leader sequence" refers to a nucleotide sequence located between the promoter sequence of a gene and the coding sequence. The translation leader sequence is present in the fully processed mRNA upstream of the translation start sequence. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency. Examples of translation leader sequences have been described (Turner and Foster (1995) *Mol. Biotechnol. 3*:225-236).

"3' Non-coding sequences" refers to nucleotide sequences located downstream of a coding sequence and includes polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht et al. (1989) *Plant Cell 1*:671-680.

"RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA (mRNA)" refers to the RNA that is without introns and that can be translated into polypeptides by the cell. "cDNA" refers to DNA that is complementary to and derived from an mRNA template. The cDNA can be single-stranded or converted to double stranded form using, for example, the Klenow fragment of DNA polymerase I. "Sense RNA" refers to an RNA transcript that includes the mRNA and can be translated into a polypeptide by the cell. "Antisense RNA" refers to an RNA transcript that is complementary to all or part of a target primary transcript or mRNA

and that blocks the expression of a target gene (see U.S. Patent No. 5,107,065, incorporated herein by reference). The complementarity of an antisense RNA may be with any part of the specific nucleotide sequence, i.e., at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. "Functional RNA" refers to sense RNA, antisense RNA, ribozyme RNA, or other RNA that may not be translated but yet has an effect on cellular processes.

The term "operably linked" refers to the association of two or more nucleic acid fragments so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term "expression", as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the invention. "Expression" may also refer to translation of mRNA into a polypeptide. "Antisense inhibition" refers to the production of antisense RNA transcripts capable of suppressing the expression of the target protein. "Overexpression" refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or non-transformed organisms. "Co-suppression" refers to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially similar foreign or endogenous genes (U.S. Patent No. 5,231,020, incorporated herein by reference).

A "protein" or "polypeptide" is a chain of amino acids arranged in a specific order determined by the coding sequence in a polynucleotide encoding the polypeptide. Each protein or polypeptide has a unique function.

"Altered levels" or "altered expression" refers to the production of gene product(s) in transgenic organisms in amounts or proportions that differ from that of normal or non-transformed organisms.

"Null mutant" refers to a host cell which either lacks the expression of a certain polypeptide or expresses a polypeptide which is inactive or does not have any detectable expected enzymatic function.

"Mature protein" refers to a post-translationally processed polypeptide; i.e., one from which any pre- or propeptides present in the primary translation product have been removed. "Precursor protein" refers to the primary product of translation of mRNA; i.e., with pre- and propeptides still present. Pre- and propeptides may be but are not limited to intracellular localization signals.

A "chloroplast transit peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the chloroplast or other plastid types present in the cell in which the protein is made. "Chloroplast transit sequence" refers to a

nucleotide sequence that encodes a chloroplast transit peptide. A "signal peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the secretory system (Chrispeels (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol. 42*:21-53). If the protein is to be directed to a vacuole, a vacuolar targeting signal (*supra*) can further be added, or if to the endoplasmic reticulum, an endoplasmic reticulum retention signal (*supra*) may be added. If the protein is to be directed to the nucleus, any signal peptide present should be removed and instead a nuclear localization signal included (Raikhel (1992) *Plant Phys. 100*:1627-1632).

"Transformation" refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic" organisms. Examples of methods of plant transformation include *Agrobacterium*-mediated transformation (De Blaere et al. (1987) *Meth. Enzymol. 143*:277) and particle-accelerated or "gene gun" transformation technology (Klein et al. (1987) *Nature (London) 327*:70-73; U.S. Patent No. 4,945,050, incorporated herein by reference). Thus, isolated polynucleotides of the present invention can be incorporated into recombinant constructs, typically DNA constructs, capable of introduction into and replication in a host cell. Such a construct can be a vector that includes a replication system and sequences that are capable of transcription and translation of a polypeptide-encoding sequence in a given host cell. A number of vectors suitable for stable transfection of plant cells or for the establishment of transgenic plants have been described in, e.g., Pouwels et al., Cloning Vectors: A Laboratory Manual, 1985, supp. 1987; Weissbach and Weissbach, Methods for Plant Molecular Biology, Academic Press, 1989; and Flevin et al., Plant Molecular Biology Manual, Kluwer Academic Publishers, 1990. Typically, plant expression vectors include, for example, one or more cloned plant genes under the transcriptional control of 5' and 3' regulatory sequences and a dominant selectable marker. Such plant expression vectors also can contain a promoter regulatory region (e.g., a regulatory region controlling inducible or constitutive, environmentally- or developmentally-regulated, or cell- or tissue-specific expression), a transcription initiation start site, a ribosome binding site, an RNA processing signal, a transcription termination site, and/or a polyadenylation signal.

Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described more fully in Sambrook et al. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1989 (hereinafter "Maniatis").

"PCR" or "polymerase chain reaction" is a technique used for the amplification of specific DNA segments (U.S. Patent Nos. 4,683,195 and 4,800,159).

The present invention concerns an isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of: (a) a first nucleotide sequence encoding a

polypeptide of at least 50 amino acids having at least 75% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, and 34, or (b) a second nucleotide sequence comprising the complement of the first nucleotide sequence.

Preferably, the first nucleotide sequence comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33, that codes for the polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, and 34.

Nucleic acid fragments encoding at least a substantial portion of several CDKI have been isolated and identified by comparison of random plant cDNA sequences to public databases containing nucleotide and protein sequences using the BLAST algorithms well known to those skilled in the art. The nucleic acid fragments of the instant invention may be used to isolate cDNAs and genes encoding homologous proteins from the same or other plant species. Isolation of homologous genes using sequence-dependent protocols is well known in the art. Examples of sequence-dependent protocols include, but are not limited to, methods of nucleic acid hybridization, and methods of DNA and RNA amplification as exemplified by various uses of nucleic acid amplification technologies (e.g., polymerase chain reaction, ligase chain reaction).

For example, genes encoding other CDKI, either as cDNAs or genomic DNAs, could be isolated directly by using all or a substantial portion of the instant nucleic acid fragments as DNA hybridization probes to screen libraries from any desired plant employing methodology well known to those skilled in the art. Specific oligonucleotide probes based upon the instant nucleic acid sequences can be designed and synthesized by methods known in the art (Maniatis). Moreover, the entire sequence(s) can be used directly to synthesize DNA probes by methods known to the skilled artisan such as random primer DNA labeling, nick translation, end-labeling techniques, or RNA probes using available *in vitro* transcription systems. In addition, specific primers can be designed and used to amplify a part or all of the instant sequences. The resulting amplification products can be labeled directly during amplification reactions or labeled after amplification reactions, and used as probes to isolate full length cDNA or genomic fragments under conditions of appropriate stringency.

In addition, two short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols to amplify longer nucleic acid fragments encoding homologous genes from DNA or RNA. The polymerase chain reaction may also be performed on a library of cloned nucleic acid fragments wherein the sequence of one primer is derived from the instant nucleic acid fragments, and the sequence of the other primer takes advantage of the presence of the polyadenylic acid tracts to the 3' end of the mRNA precursor encoding plant genes. Alternatively, the second primer sequence may be based

upon sequences derived from the cloning vector. For example, the skilled artisan can follow the RACE protocol (Frohman et al. (1988) *Proc. Natl. Acad. Sci. USA 85*:8998-9002) to generate cDNAs by using PCR to amplify copies of the region between a single point in the transcript and the 3' or 5' end. Primers oriented in the 3' and 5' directions can be designed from the instant sequences. Using commercially available 3' RACE or 5' RACE systems (BRL), specific 3' or 5' cDNA fragments can be isolated (Ohara et al. (1989) *Proc. Natl. Acad. Sci. USA 86*:5673-5677; Loh et al. (1989) *Science 243*:217-220). Products generated by the 3' and 5' RACE procedures can be combined to generate full-length cDNAs (Frohman and Martin (1989) *Techniques 1*:165). Consequently, a polynucleotide comprising a nucleotide sequence of at least one of 60 (preferably one of at least 40, most preferably one of at least 30) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33 and the complement of such nucleotide sequences may be used in such methods to obtain a nucleic acid fragment encoding a substantial portion of an amino acid sequence of a polypeptide.

The present invention relates to a method of obtaining a nucleic acid fragment encoding a substantial portion of a CDKI polypeptide, preferably a substantial portion of a plant CDKI polypeptide, comprising the steps of: synthesizing an oligonucleotide primer comprising a nucleotide sequence of at least one of 60 (preferably at least one of 40, most preferably at least one of 30) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33, and the complement of such nucleotide sequences; and amplifying a nucleic acid fragment (preferably a cDNA inserted in a cloning vector) using the oligonucleotide primer. The amplified nucleic acid fragment preferably will encode a substantial portion of a CDKI polypeptide.

Availability of the instant nucleotide and deduced amino acid sequences facilitates immunological screening of cDNA expression libraries. Synthetic peptides representing substantial portions of the instant amino acid sequences may be synthesized. These peptides can be used to immunize animals to produce polyclonal or monoclonal antibodies with specificity for peptides or proteins comprising the amino acid sequences. These antibodies can be then be used to screen cDNA expression libraries to isolate full-length cDNA clones of interest (Lerner (1984) *Adv. Immunol. 36*:1-34; Maniatis).

In another embodiment, this invention concerns viruses and host cells comprising either the chimeric genes of the invention as described herein or an isolated polynucleotide of the invention as described herein. Examples of host cells which can be used to practice the invention include, but are not limited to, yeast, bacteria, and plants.

As was noted above, the nucleic acid fragments of the instant invention may be used to create transgenic plants in which the disclosed polypeptides are present at higher or lower

levels than normal or in cell types or developmental stages in which they are not normally found. This would have the effect of altering growth rate or viability of the plants.

Overexpression of the proteins of the instant invention may be accomplished by first constructing a chimeric gene in which the coding region is operably linked to a promoter capable of directing expression of a gene in the desired tissues at the desired stage of development. The chimeric gene may comprise promoter sequences and translation leader sequences derived from the same genes. 3' Non-coding sequences encoding transcription termination signals may also be provided. The instant chimeric gene may also comprise one or more introns in order to facilitate gene expression.

Plasmid vectors comprising the instant isolated polynucleotide (or chimeric gene) may be constructed. The choice of plasmid vector is dependent upon the method that will be used to transform host plants. The skilled artisan is well aware of the genetic elements that must be present on the plasmid vector in order to successfully transform, select and propagate host cells containing the chimeric gene. The skilled artisan will also recognize that different independent transformation events will result in different levels and patterns of expression (Jones et al. (1985) *EMBO J. 4*:2411-2418; De Almeida et al. (1989) *Mol. Gen. Genetics 218*:78-86), and thus that multiple events must be screened in order to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished by Southern analysis of DNA, Northern analysis of mRNA expression, Western analysis of protein expression, or phenotypic analysis.

For some applications it may be useful to direct the instant polypeptides to different cellular compartments, or to facilitate secretion from the cell. It is thus envisioned that the chimeric gene described above may be further supplemented by directing the coding sequence to encode the instant polypeptides with appropriate intracellular targeting sequences such as transit sequences (Keegstra (1989) *Cell 56*:247-253), signal sequences or sequences encoding endoplasmic reticulum localization (Chrispeels (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol. 42*:21-53), or nuclear localization signals (Raikhel (1992) *Plant Phys.100*:1627-1632) with or without removing targeting sequences that are already present. While the references cited give examples of each of these, the list is not exhaustive and more targeting signals of use may be discovered in the future.

It may also be desirable to reduce or eliminate expression of genes encoding the instant polypeptides in plants for some applications. In order to accomplish this, a chimeric gene designed for co-suppression of the instant polypeptide can be constructed by linking a gene or gene fragment encoding that polypeptide to plant promoter sequences.

Alternatively, a chimeric gene designed to express antisense RNA for all or part of the instant nucleic acid fragment can be constructed by linking the gene or gene fragment in reverse orientation to plant promoter sequences. Either the cosuppression or antisense

chimeric genes could be introduced into plants via transformation wherein expression of the corresponding endogenous genes are reduced or eliminated.

Molecular genetic solutions to the generation of plants with altered gene expression have a decided advantage over more traditional plant breeding approaches. Changes in plant phenotypes can be produced by specifically inhibiting expression of one or more genes by antisense inhibition or cosuppression (U.S. Patent Nos. 5,190,931, 5,107,065 and 5,283,323). An antisense or cosuppression construct would act as a dominant negative regulator of gene activity. While conventional mutations can yield negative regulation of gene activity these effects are most likely recessive. The dominant negative regulation available with a transgenic approach may be advantageous from a breeding perspective. In addition, the ability to restrict the expression of specific phenotype to the reproductive tissues of the plant by the use of tissue specific promoters may confer agronomic advantages relative to conventional mutations which may have an effect in all tissues in which a mutant gene is ordinarily expressed.

The person skilled in the art will know that special considerations are associated with the use of antisense or cosuppression technologies in order to reduce expression of particular genes. For example, the proper level of expression of sense or antisense genes may require the use of different chimeric genes utilizing different regulatory elements known to the skilled artisan. Once transgenic plants are obtained by one of the methods described above, it will be necessary to screen individual transgenics for those that most effectively display the desired phenotype. Accordingly, the skilled artisan will develop methods for screening large numbers of transformants. The nature of these screens will generally be chosen on practical grounds. For example, one can screen by looking for changes in gene expression by using antibodies specific for the protein encoded by the gene being suppressed, or one could establish assays that specifically measure enzyme activity. A preferred method will be one which allows large numbers of samples to be processed rapidly, since it will be expected that a large number of transformants will be negative for the desired phenotype.

In another embodiment, the present invention concerns a polypeptide of at least 50 amino acids that has at least 75% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, and 34.

The instant polypeptides (or substantial portions thereof) may be produced in heterologous host cells, particularly in the cells of microbial hosts, and can be used to prepare antibodies to the proteins by methods well known to those skilled in the art. The antibodies are useful for detecting the polypeptides of the instant invention *in situ* in cells or *in vitro* in cell extracts. Preferred heterologous host cells for production of the instant polypeptides are microbial hosts. Microbial expression systems and expression vectors containing regulatory sequences that direct high level expression of foreign proteins are well

known to those skilled in the art. Any of these could be used to construct a chimeric gene for production of the instant polypeptides. This chimeric gene could then be introduced into appropriate microorganisms via transformation to provide high level expression of the encoded CDKI. An example of a vector for high level expression of the instant polypeptides in a bacterial host is provided (Example 6).

Additionally, the instant polypeptides can be used as a target to facilitate design and/or for identification of inhibitors of those enzymes that may be useful as herbicides. This is desirable because the polypeptides described herein regulate key components of the plant cell cycle. Accordingly, inhibition of the activity of one or more of the peptides described herein could lead to inhibition of plant growth. Thus, the instant polypeptides could be appropriate for new herbicide discovery and design.

All or a substantial portion of the polynucleotides of the instant invention may also be used as probes for genetically and physically mapping the genes that they are a part of, and used as markers for traits linked to those genes. Such information may be useful in plant breeding in order to develop lines with desired phenotypes. For example, the instant nucleic acid fragments may be used as restriction fragment length polymorphism (RFLP) markers. Southern blots (Maniatis) of restriction-digested plant genomic DNA may be probed with the nucleic acid fragments of the instant invention. The resulting banding patterns may then be subjected to genetic analyses using computer programs such as MapMaker (Lander et al. (1987) *Genomics 1*:174-181) in order to construct a genetic map. In addition, the nucleic acid fragments of the instant invention may be used to probe Southern blots containing restriction endonuclease-treated genomic DNAs of a set of individuals representing parent and progeny of a defined genetic cross. Segregation of the DNA polymorphisms is noted and used to calculate the position of the instant nucleic acid sequence in the genetic map previously obtained using this population (Botstein et al. (1980) *Am. J. Hum. Genet. 32*:314-331).

The production and use of plant gene-derived probes for use in genetic mapping is described in Bernatzky and Tanksley (1986) *Plant Mol. Biol. Reporter 4*:37-41. Numerous publications describe genetic mapping of specific cDNA clones using the methodology outlined above or variations thereof. For example, F2 intercross populations, backcross populations, randomly mated populations, near isogenic lines, and other sets of individuals may be used for mapping. Such methodologies are well known to those skilled in the art.

Nucleic acid probes derived from the instant nucleic acid sequences may also be used for physical mapping (i.e., placement of sequences on physical maps; *see* Hoheisel et al. In: *Nonmammalian Genomic Analysis: A Practical Guide*, Academic press 1996, pp. 319-346, and references cited therein).

In another embodiment, nucleic acid probes derived from the instant nucleic acid sequences may be used in direct fluorescence *in situ* hybridization (FISH) mapping (Trask

(1991) *Trends Genet.* 7:149-154). Although current methods of FISH mapping favor use of large clones (several to several hundred KB; see Laan et al. (1995) *Genome Res.* 5:13-20), improvements in sensitivity may allow performance of FISH mapping using shorter probes.

A variety of nucleic acid amplification-based methods of genetic and physical mapping may be carried out using the instant nucleic acid sequences. Examples include allele-specific amplification (Kazazian (1989) *J. Lab. Clin. Med.* 11:95-96), polymorphism of PCR-amplified fragments (CAPS; Sheffield et al. (1993) *Genomics* 16:325-332), allele-specific ligation (Landegren et al. (1988) *Science* 241:1077-1080), nucleotide extension reactions (Sokolov (1990) *Nucleic Acid Res.* 18:3671), Radiation Hybrid Mapping (Walter et al. (1997) *Nat. Genet.* 7:22-28) and Happy Mapping (Dear and Cook (1989) *Nucleic Acid Res.* 17:6795-6807). For these methods, the sequence of a nucleic acid fragment is used to design and produce primer pairs for use in the amplification reaction or in primer extension reactions. The design of such primers is well known to those skilled in the art. In methods employing PCR-based genetic mapping, it may be necessary to identify DNA sequence differences between the parents of the mapping cross in the region corresponding to the instant nucleic acid sequence. This, however, is generally not necessary for mapping methods.

Loss of function mutant phenotypes may be identified for the instant cDNA clones either by targeted gene disruption protocols or by identifying specific mutants for these genes contained in a maize population carrying mutations in all possible genes (Ballinger and Benzer (1989) *Proc. Natl. Acad. Sci USA* 86:9402-9406; Koes et al. (1995) *Proc. Natl. Acad. Sci USA* 92:8149-8153; Bensen et al. (1995) *Plant Cell* 7:75-84). The latter approach may be accomplished in two ways. First, short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols in conjunction with a mutation tag sequence primer on DNAs prepared from a population of plants in which Mutator transposons or some other mutation-causing DNA element has been introduced (see Bensen, *supra*). The amplification of a specific DNA fragment with these primers indicates the insertion of the mutation tag element in or near the plant gene encoding the instant polypeptides. Alternatively, the instant nucleic acid fragment may be used as a hybridization probe against PCR amplification products generated from the mutation population using the mutation tag sequence primer in conjunction with an arbitrary genomic site primer, such as that for a restriction enzyme site-anchored synthetic adaptor. With either method, a plant containing a mutation in the endogenous gene encoding the instant polypeptides can be identified and obtained. This mutant plant can then be used to determine or confirm the natural function of the instant polypeptides disclosed herein.

EXAMPLES

The present invention is further defined in the following Examples, in which parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be

understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions. Thus, various modifications of the invention in addition to those shown and described herein will be apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the appended claims.

The disclosure of each reference set forth herein is incorporated herein by reference in its entirety.

EXAMPLE 1

Composition of cDNA Libraries; Isolation and Sequencing of cDNA Clones

cDNA libraries representing mRNAs from various corn, rice, soybean, and wheat tissues were prepared. The characteristics of the libraries are described below.

TABLE 2

cDNA Libraries from Corn, Rice, Soybean, and Wheat

Library	Tissue	Clone
cbn2n	Corn Developing Kernel Two Days After Pollination*	cbn2n.pk0010.h7
cen3n	Corn Endosperm 20 Days After Pollination*	cen3n.pk0013.cp, cen3n.pk0115.a6, cen3n.pk0151.d2
crln	Corn Root From 7 Day Old Seedlings*	crln.pk0052.f11, crln.pk0195.b3
csl	Corn Leaf Sheath From 5 Week Old Plant	csl.pk0068.c12
csiln	Corn Silk*	csiln.pk0050.e6, csiln.pk0050.e6:fis
ctaln	Corn Tassel*	ctaln.pk0070.d4
p0023	Leaf: Gene M1C07 (leucine-rich repeat), family 3-B7, induces resistance prior to genetic lesion formation. about one month after planting in green house	p0023.clrag76r
p0058	Honey N pearl (sweet corn hybrid) shoot culture. It was initiated on 2/28/96 from seed derived meristems. The culture was maintained on 273N medium	p0058.chpbm23rb
p0086	P0067 screened 1; 11 DAP pericarp	p0086.cbsaj18f, p0086.cbsaj18r, p0086.cbsaj18rb
p0095	Ear leaf sheath, screened 1 Growth conditions: field; control or untreated tissues Growth stage: 2-3 weeks after pollen shed; plants were allowed to pollinate naturally	p0095.cwsbc53r
rds2c	Rice (*Oryza sativa*, YM) developing seeds in the middle of the plant	rds2c.pk008.o24

Library	Tissue	Clone
rsr9n	Rice (*Oryza sativa* L.) leaf (15 DAG) following infection of *Magnaporta grisea* (4360-R-62 and 4360-R-67) from 2 to 72 hrs*	rsr9n.pk003.g12, rsr9n.pk003.g12:fis
sdp4c	Soybean (*Glycine max* L.) developing pods 10-12 mm	sdp4c.pk025.k23
sl2	Soybean Two-Week-Old Developing Seedlings Treated With 2.5 ppm chlorimuron	sl2.pk0008.d2, sl2.pk0008.d2:fis, sl2.pk0117.h4
sls1c	Soybean (*Glycine max* L., S1990) infected with *Sclerotinia sclerotiorum* mycelium	sls1c.pk007.h20
wre1n	Wheat Root From 7 Day Old Etiolated Seedling*	wre1n.pk0031.g2, wre1n.pk0031.g2:fis

*These libraries were normalized essentially as described in U.S. Patent No. 5,482,845, incorporated herein by reference.

cDNA libraries may be prepared by any one of many methods available. For example, the cDNAs may be introduced into plasmid vectors by first preparing the cDNA libraries in Uni-ZAP™ XR vectors according to the manufacturer's protocol (Stratagene Cloning Systems, La Jolla, CA). The Uni-ZAP™ XR libraries are converted into plasmid libraries according to the protocol provided by Stratagene. Upon conversion, cDNA inserts will be contained in the plasmid vector pBluescript. In addition, the cDNAs may be introduced directly into precut Bluescript II SK(+) vectors (Stratagene) using T4 DNA ligase (New England Biolabs), followed by transfection into DH10B cells according to the manufacturer's protocol (GIBCO BRL Products). Once the cDNA inserts are in plasmid vectors, plasmid DNAs are prepared from randomly picked bacterial colonies containing recombinant pBluescript plasmids, or the insert cDNA sequences are amplified via polymerase chain reaction using primers specific for vector sequences flanking the inserted cDNA sequences. Amplified insert DNAs or plasmid DNAs are sequenced in dye-primer sequencing reactions to generate partial cDNA sequences (expressed sequence tags or "ESTs"; see Adams et al., (1991) *Science 252*:1651-1656). The resulting ESTs are analyzed using a Perkin Elmer Model 377 fluorescent sequencer.

EXAMPLE 2

Identification of cDNA Clones

cDNA clones encoding CDKI were identified by conducting BLAST (Basic Local Alignment Search Tool; Altschul et al. (1993) *J. Mol. Biol. 215*:403-410; see also www.ncbi.nlm.nih.gov/BLAST/) searches for similarity to sequences contained in the BLAST "nr" database (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the last major release of the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The

cDNA sequences obtained in Example 1 were analyzed for similarity to all publicly available DNA sequences contained in the "nr" database using the BLASTN algorithm provided by the National Center for Biotechnology Information (NCBI). The DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database using the BLASTX algorithm (Gish and States (1993) *Nat. Genet. 3*:266-272) provided by the NCBI. For convenience, the P-value (probability) of observing a match of a cDNA sequence to a sequence contained in the searched databases merely by chance as calculated by BLAST are reported herein as "pLog" values, which represent the negative of the logarithm of the reported P-value. Accordingly, the greater the pLog value, the greater the likelihood that the cDNA sequence and the BLAST "hit" represent homologous proteins.

EXAMPLE 3

Characterization of cDNA Clones Encoding Cyclin Dependent Kinase Inhibitors

The BLASTX search using the EST sequences from clones listed in Table 3 revealed similarity of the polypeptides encoded by the cDNAs to CDKIs from the flowering short-day plant *Chenopodium* [*Chenopodium rubrum*] (NCBI Accession No. gi 2653281), the nematode [*Caenorhabditis elegans*] (NCBI Accession No. gi 2731583), and the flowering weed *Arabidopsis* [*Arabidopsis thaliana*] (NCBI Accession No. gi 2914702). Shown in Table 3 are the BLAST results for individual ESTs ("EST"), the sequences of the entire cDNA inserts comprising the indicated cDNA clones ("FIS"), contigs assembled from two or more ESTs ("Contig"), contigs assembled from an FIS and one or more ESTs ("Contig*"), or sequences encoding the entire protein derived from an FIS, a contig, or an FIS and PCR ("CGS"):

TABLE 3

BLAST Results for Sequences Encoding Polypeptides Homologous to Cyclin Dependent Kinase Inhibitors

Clone	Status	BLAST pLog Score 2653281	2914702	2731583
csi1n.pk0050.e6	EST	6.4		
rsr9n.pk003.g12	EST		18.7	
sl2.pk0008.d2	EST	13.0		
wre1n.pk0031.g2	EST		19.4	
csi1n.pk0050.e6:fis	FIS	8.2		
contig of: cr1n.pk0052.f11, cr1n.pk0195.b3, p0095.cwsbc53r	Contig	6.7		
p0023.clrag76r	EST			1.9

Clone	Status	BLAST pLog Score 2653281	2914702	2731583
contig of: cen3n.pk0013.c9, cen3n.pk0115.a6, cen3n.pk0151.d2, cs1.pk0068.c12, cta1n.pk0070.d4, p0058.chpbm23rb	Contig		16.5	
contig of: cbn2n.pk0010.h7, p0086.cbsaj18f, p0086.cbsaj18r, p0086.cbsaj18rb	Contig	7.0		
rsr9n.pk003.g12:fis	FIS		13.7	
rds2c.pk008.o24	EST	3.15		
sl2.pk0008.d2:fis	FIS	14.2		
sdp4c.pk025.k23	CGS	20.2		
sl2.pk0117.h4	EST	15.1		
sls1c.pk007.h20-5'	EST		4.2	
sls1c.pk007.h20-3'	EST	13.7		
wre1n.pk0031.g2:fis	FIS		22.3	

The first four sequences (csi1n.pk0050.e6, rsr9n.pk003.g12, sl2.pk0008.d2, and wre1n.pk0031.g2) represent the sequences presented in the priority filing of this application (U.S. Provisional Application No. 60/128,192). The ":fis" sequences are the complete insert seequences of those cDNA clones. The sls1c.pk007.h20 entries are the 5' and 3' end sequences (respectively) of a cDNA clone for a full-length soybean CDKI. In total, the table contains sequences representing distinct CDKI sequences from four corn genes, two rice genes, four soybean genes, and one wheat gene.

Figure 1 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, and 34 compared to the amino acid sequences of the closest art, the short-day flowering plant *Chenopodium rubrum* [NCBI Accession No. gi 265281, SEQ ID NO:35], the nematode *Caenorhabditis elegans* [NCBI Accession No. gi 2731583, SEQ ID NO:36], and the flowering weed *Arabidopsis thaliana* [NCBI Accession No. gi 2914702, SEQ ID NO:39]. The data in Table 4 represents a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, and 34 compared to the amino acid sequences of the closest art, the sort-day flowering plant *Chenopodium rubrum* [NCBI Accession No. gi 265281, SEQ ID NO:35], the nematode *Caenorhabditis elegans* [NCBI

Accession No. gi 2731583, SEQ ID NO:36], or the flowering weed *Arabidopsis thaliana* [NCBI Accession No. gi 2914702, SEQ ID NO:39].

TABLE 4

Percent Identity of Amino Acid Sequences Deduced From the Nucleotide Sequences of cDNA Clones Encoding Polypeptides Homologous to CDKIs

SEQ ID NO.	% Identity to 2653281	% Identity to 2914702	% Identity to 2731583
2	**45.3**	39.6	18.9
4	45.7	**56.5**	17.4
6	**71.1**	47.4	18.4
8	50.0	**65.2**	19.6
10	**27.0**	24.6	13.5
12	**26.4**	23.2	16.0
14	13.9	13.9	**17.4**
16	17.9	**28.9**	11.1
18	**38.9**	27.8	18.5
20	27.3	**42.4**	17.2
22	**37.8**	43.2	16.2
24	**41.4**	28.7	16.1
26	**28.6**	19.0	14.6
28	**50.0**	38.3	18.3
30	13.8	**37.9**	13.8
32	**38.6**	37.5	15.9
34	19.5	**32.1**	15.3

Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS. 5*:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. The underlined and bold percentages correspond to the highest homology BLAST sequence shown in Table 3. Sequence alignments, BLAST scores and probabilities indicate that the nucleic acid fragments comprising the instant cDNA clones encode a substantial portion of a plant CDKI.

EXAMPLE 4

Expression of Chimeric Genes in Monocot Cells

A chimeric gene comprising a cDNA encoding the instant polypeptides in sense orientation with respect to the maize 27 kD zein promoter that is located 5' to the cDNA fragment, and the 10 kD zein 3' end that is located 3' to the cDNA fragment, can be constructed. The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites (NcoI or SmaI) can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the digested vector pML103 as described below. Amplification is then performed in a standard PCR. The amplified DNA is then digested with restriction enzymes NcoI and SmaI and fractionated on an agarose gel. The appropriate band can be isolated from the gel and combined with a 4.9 kb NcoI-SmaI fragment of the plasmid pML103. Plasmid pML103 has been deposited under the terms of the Budapest Treaty at ATCC (American Type Culture Collection, 10801 University Blvd., Manassas, VA 20110-2209), and bears accession number ATCC 97366. The DNA segment from pML103 contains a 1.05 kb SalI-NcoI promoter fragment of the maize 27 kD zein gene and a 0.96 kb SmaI-SalI fragment from the 3' end of the maize 10 kD zein gene in the vector pGem9Zf(+) (Promega). Vector and insert DNA can be ligated at 15°C overnight, essentially as described (Maniatis). The ligated DNA may then be used to transform *E. coli* XL1-Blue (Epicurian Coli XL-1 Blue™; Stratagene). Bacterial transformants can be screened by restriction enzyme digestion of plasmid DNA and limited nucleotide sequence analysis using the dideoxy chain termination method (Sequenase™ DNA Sequencing Kit; U.S. Biochemical). The resulting plasmid construct would comprise a chimeric gene encoding, in the 5' to 3' direction, the maize 27 kD zein promoter, a cDNA fragment encoding the instant polypeptides, and the 10 kD zein 3' region.

The chimeric gene described above can then be introduced into corn cells by the following procedure. Immature corn embryos can be dissected from developing caryopses derived from crosses of the inbred corn lines H99 and LH132. The embryos are isolated 10 to 11 days after pollination when they are 1.0 to 1.5 mm long. The embryos are then placed with the axis-side facing down and in contact with agarose-solidified N6 medium (Chu et al. (1975) *Sci. Sin. Peking* 18:659-668). The embryos are kept in the dark at 27°C. Friable embryogenic callus consisting of undifferentiated masses of cells with somatic proembryoids and embryoids borne on suspensor structures proliferates from the scutellum of these immature embryos. The embryogenic callus isolated from the primary explant can be cultured on N6 medium and sub-cultured on this medium every 2 to 3 weeks.

The plasmid, p35S/Ac (obtained from Dr. Peter Eckes, Hoechst Ag, Frankfurt, Germany) may be used in transformation experiments in order to provide for a selectable marker. This plasmid contains the *Pat* gene (see European Patent Publication 0 242 236)

which encodes phosphinothricin acetyl transferase (PAT). The enzyme PAT confers resistance to herbicidal glutamine synthetase inhibitors such as phosphinothricin. The *pat* gene in p35S/Ac is under the control of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*.

The particle bombardment method (Klein et al. (1987) *Nature* 327:70-73) may be used to transfer genes to the callus culture cells. According to this method, gold particles (1 μm in diameter) are coated with DNA using the following technique. Ten μg of plasmid DNAs are added to 50 μL of a suspension of gold particles (60 mg per mL). Calcium chloride (50 μL of a 2.5 M solution) and spermidine free base (20 μL of a 1.0 M solution) are added to the particles. The suspension is vortexed during the addition of these solutions. After 10 minutes, the tubes are briefly centrifuged (5 sec at 15,000 rpm) and the supernatant removed. The particles are resuspended in 200 μL of absolute ethanol, centrifuged again and the supernatant removed. The ethanol rinse is performed again and the particles resuspended in a final volume of 30 μL of ethanol. An aliquot (5 μL) of the DNA-coated gold particles can be placed in the center of a Kapton™ flying disc (Bio-Rad Labs). The particles are then accelerated into the corn tissue with a Biolistic™ PDS-1000/He (Bio-Rad Instruments, Hercules CA), using a helium pressure of 1000 psi, a gap distance of 0.5 cm and a flying distance of 1.0 cm.

For bombardment, the embryogenic tissue is placed on filter paper over agarose-solidified N6 medium. The tissue is arranged as a thin lawn and covered a circular area of about 5 cm in diameter. The petri dish containing the tissue can be placed in the chamber of the PDS-1000/He approximately 8 cm from the stopping screen. The air in the chamber is then evacuated to a vacuum of 28 inches of Hg. The macrocarrier is accelerated with a helium shock wave using a rupture membrane that bursts when the He pressure in the shock tube reaches 1000 psi.

Seven days after bombardment the tissue can be transferred to N6 medium that contains gluphosinate (2 mg per liter) and lacks casein or proline. The tissue continues to grow slowly on this medium. After an additional 2 weeks the tissue can be transferred to fresh N6 medium containing gluphosinate. After 6 weeks, areas of about 1 cm in diameter of actively growing callus can be identified on some of the plates containing the glufosinate-supplemented medium. These calli may continue to grow when sub-cultured on the selective medium.

Plants can be regenerated from the transgenic callus by first transferring clusters of tissue to N6 medium supplemented with 0.2 mg per liter of 2,4-D. After two weeks the tissue can be transferred to regeneration medium (Fromm et al. (1990) *Bio/Technology* 8:833-839).

EXAMPLE 5

Expression of Chimeric Genes in Dicot Cells

A seed-specific construct composed of the promoter and transcription terminator from the gene encoding the β subunit of the seed storage protein phaseolin from the bean *Phaseolus vulgaris* (Doyle et al. (1986) *J. Biol. Chem. 261*:9228-9238) can be used for expression of the instant polypeptides in transformed soybean. The phaseolin construct includes about 500 nucleotides upstream (5') from the translation initiation codon and about 1650 nucleotides downstream (3') from the translation stop codon of phaseolin. Between the 5' and 3' regions are the unique restriction endonuclease sites Nco I (which includes the ATG translation initiation codon), Sma I, Kpn I and Xba I. The entire constructcassette is flanked by Hind III sites.

The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the expression vector. Amplification is then performed as described above, and the isolated fragment is inserted into a pUC18 vector carrying the seed construct.

Soybean embryos may then be transformed with the expression vector comprising sequences encoding the instant polypeptides. To induce somatic embryos, cotyledons, 3-5 mm in length dissected from surface sterilized, immature seeds of the soybean cultivar A2872, can be cultured in the light or dark at 26°C on an appropriate agar medium for 6-10 weeks. Somatic embryos which produce secondary embryos are then excised and placed into a suitable liquid medium. After repeated selection for clusters of somatic embryos which multiplied as early, globular staged embryos, the suspensions are maintained as described below.

Soybean embryogenic suspension cultures can be maintained in 35 mL liquid media on a rotary shaker, 150 rpm, at 26°C with florescent lights on a 16:8 hour day/night schedule. Cultures are subcultured every two weeks by inoculating approximately 35 mg of tissue into 35 mL of liquid medium.

Soybean embryogenic suspension cultures may then be transformed by the method of particle gun bombardment (Klein et al. (1987) *Nature* (London) *327*:70-73, U.S. Patent No. 4,945,050). A DuPont Biolistic™ PDS1000/HE instrument (helium retrofit) can be used for these transformations.

A selectable marker gene which can be used to facilitate soybean transformation is a chimeric gene composed of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature 313*:810-812), the hygromycin phosphotransferase gene from plasmid pJR225 (from *E. coli*; Gritz et al.(1983) *Gene 25*:179-188) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*. The seed construct comprising the phaseolin 5' region, the fragment encoding the instant polypeptides and the

phaseolin 3' region can be isolated as a restriction fragment. This fragment can then be inserted into a unique restriction site of the vector carrying the marker gene.

To 50 µL of a 60 mg/mL 1 µm gold particle suspension is added (in order): 5 µL DNA (1 µg/µL), 20 µL spermidine (0.1 M), and 50 µL $CaCl_2$ (2.5 M). The particle preparation is then agitated for three minutes, spun in a microfuge for 10 seconds and the supernatant removed. The DNA-coated particles are then washed once in 400 µL 70% ethanol and resuspended in 40 µL of anhydrous ethanol. The DNA/particle suspension can be sonicated three times for one second each. Five µL of the DNA-coated gold particles are then loaded on each macro carrier disk.

Approximately 300-400 mg of a two-week-old suspension culture is placed in an empty 60x15 mm petri dish and the residual liquid removed from the tissue with a pipette. For each transformation experiment, approximately 5-10 plates of tissue are normally bombarded. Membrane rupture pressure is set at 1100 psi and the chamber is evacuated to a vacuum of 28 inches mercury. The tissue is placed approximately 3.5 inches away from the retaining screen and bombarded three times. Following bombardment, the tissue can be divided in half and placed back into liquid and cultured as described above.

Five to seven days post bombardment, the liquid media may be exchanged with fresh media, and eleven to twelve days post bombardment with fresh media containing 50 mg/mL hygromycin. This selective media can be refreshed weekly. Seven to eight weeks post bombardment, green, transformed tissue may be observed growing from untransformed, necrotic embryogenic clusters. Isolated green tissue is removed and inoculated into individual flasks to generate new, clonally propagated, transformed embryogenic suspension cultures. Each new line may be treated as an independent transformation event. These suspensions can then be subcultured and maintained as clusters of immature embryos or regenerated into whole plants by maturation and germination of individual somatic embryos.

EXAMPLE 6

Expression of Chimeric Genes in Microbial Cells

The cDNAs encoding the instant polypeptides can be inserted into the T7 *E. coli* expression vector pBT430. This vector is a derivative of pET-3a (Rosenberg et al. (1987) *Gene 56*:125-135) which employs the bacteriophage T7 RNA polymerase/T7 promoter system. Plasmid pBT430 was constructed by first destroying the EcoR I and Hind III sites in pET-3a at their original positions. An oligonucleotide adaptor containing EcoR I and Hind III sites was inserted at the BamH I site of pET-3a. This created pET-3aM with additional unique cloning sites for insertion of genes into the expression vector. Then, the Nde I site at the position of translation initiation was converted to an Nco I site using oligonucleotide-directed mutagenesis. The DNA sequence of pET-3aM in this region, 5'-CATATGG, was converted to 5'-CCCATGG in pBT430.

Plasmid DNA containing a cDNA may be appropriately digested to release a nucleic acid fragment encoding the protein. This fragment may then be purified on a 1% NuSieve GTG™ low melting agarose gel (FMC). Buffer and agarose contain 10 μg/mL ethidium bromide for visualization of the DNA fragment. The fragment can then be purified from the agarose gel by digestion with GELase™ (Epicentre Technologies) according to the manufacturer's instructions, ethanol precipitated, dried and resuspended in 20 μL of water. Appropriate oligonucleotide adapters may be ligated to the fragment using T4 DNA ligase (New England Biolabs, Beverly, MA). The fragment containing the ligated adapters can be purified from the excess adapters using low melting agarose as described above. The vector pBT430 is digested, dephosphorylated with alkaline phosphatase (NEB) and deproteinized with phenol/chloroform as described above. The prepared vector pBT430 and fragment can then be ligated at 16°C for 15 hours followed by transformation into DH5 electrocompetent cells (GIBCO BRL). Transformants can be selected on agar plates containing LB media and 100 μg/mL ampicillin. Transformants containing the gene encoding the instant polypeptides are then screened for the correct orientation with respect to the T7 promoter by restriction enzyme analysis.

For high level expression, a plasmid clone with the cDNA insert in the correct orientation relative to the T7 promoter can be transformed into *E. coli* strain BL21(DE3) (Studier et al. (1986) *J. Mol. Biol. 189*:113-130). Cultures are grown in LB medium containing ampicillin (100 mg/L) at 25°C. At an optical density at 600 nm of approximately 1, IPTG (isopropylthio-β-galactoside, the inducer) can be added to a final concentration of 0.4 mM and incubation can be continued for 3 h at 25°. Cells are then harvested by centrifugation and re-suspended in 50 μL of 50 mM Tris-HCl at pH 8.0 containing 0.1 mM DTT and 0.2 mM phenyl methylsulfonyl fluoride. A small amount of 1 mm glass beads can be added and the mixture sonicated 3 times for about 5 seconds each time with a microprobe sonicator. The mixture is centrifuged and the protein concentration of the supernatant determined. One μg of protein from the soluble fraction of the culture can be separated by SDS-polyacrylamide gel electrophoresis. Gels can be observed for protein bands migrating at the expected molecular weight.

EXAMPLE 7

Evaluating Compounds for Their Ability to Inhibit the Activity of Cyclin Dependent Kinase Inhibitor

The polypeptides described herein may be produced using any number of methods known to those skilled in the art. Such methods include, but are not limited to, expression in bacteria as described in Example 6, or expression in eukaryotic cell culture, *in planta*, and using viral expression systems in suitably infected organisms or cell lines. The instant polypeptides may be expressed either as mature forms of the proteins as observed *in vivo* or as fusion proteins by covalent attachment to a variety of enzymes, proteins or affinity tags.

Common fusion protein partners include glutathione S-transferase ("GST"), thioredoxin ("Trx"), maltose binding protein, and C- and/or N-terminal hexahistidine polypeptide ("$(His)_6$"). The fusion proteins may be engineered with a protease recognition site at the fusion point so that fusion partners can be separated by protease digestion to yield intact mature enzyme. Examples of such proteases include thrombin, enterokinase and factor Xa. However, any protease can be used which specifically cleaves the peptide connecting the fusion protein and the enzyme.

Purification of the instant polypeptides, if desired, may utilize any number of separation technologies familiar to those skilled in the art of protein purification. Examples of such methods include, but are not limited to, homogenization, filtration, centrifugation, heat denaturation, ammonium sulfate precipitation, desalting, pH precipitation, ion exchange chromatography, hydrophobic interaction chromatography and affinity chromatography, wherein the affinity ligand represents a substrate, substrate analog or inhibitor. When the instant polypeptides are expressed as fusion proteins, the purification protocol may include the use of an affinity resin which is specific for the fusion protein tag attached to the expressed enzyme or an affinity resin containing ligands which are specific for the enzyme. For example, the instant polypeptides may be expressed as a fusion protein coupled to the C-terminus of thioredoxin. In addition, a $(His)_6$ peptide may be engineered into the N-terminus of the fused thioredoxin moiety to afford additional opportunities for affinity purification. Other suitable affinity resins could be synthesized by linking the appropriate ligands to any suitable resin such as Sepharose-4B. In an alternate embodiment, a thioredoxin fusion protein may be eluted using dithiothreitol; however, elution may be accomplished using other reagents which interact to displace the thioredoxin from the resin. These reagents include β–mercaptoethanol or other reduced thiol. The eluted fusion protein may be subjected to further purification by traditional means as stated above, if desired. Proteolytic cleavage of the thioredoxin fusion protein and the enzyme may be accomplished after the fusion protein is purified or while the protein is still bound to the ThioBond™ affinity resin or other resin.

Crude, partially purified or purified enzyme, either alone or as a fusion protein, may be utilized in assays for the evaluation of compounds for their ability to inhibit enzymatic activation of the instant polypeptides disclosed herein. Assays may be conducted under well known experimental conditions which permit optimal enzymatic activity. For example, assays for CDKIs are presented by Wang H., et al. (1998) *Plant J 15*:501-510; and Schuppler U., et al. (1998) *Plant Physiol 117*:667-678.

CLAIMS

What is claimed is:

1. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:

(a) a first nucleotide sequence encoding a polypeptide of at least 50 amino acids that has at least 75% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, and 34, or

(b) a second nucleotide sequence comprising a complement of the first nucleotide sequence.

2. The isolated polynucleotide of Claim 1, wherein the first nucleotide sequence comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33 that codes for the polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, and 34.

3. The isolated polynucleotide of Claim 1 wherein the nucleotide sequences are DNA.

4. The isolated polynucleotide of Claim 1 wherein the nucleotide sequences are RNA.

5. A chimeric gene comprising the isolated polynucleotide of Claim 1 operably linked to at least one suitable regulatory sequence.

6. A host cell comprising the chimeric gene of Claim 5.

7. A host cell comprising the isolated polynucleotide of Claim 1.

8. The host cell of Claim 7 wherein the host cell is selected from the group consisting of yeast, bacteria, and plant.

9. A virus comprising the isolated polynucleotide of Claim 1.

10. A polypeptide of at least 50 amino acids that has at least 75% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, and 34.

11. A method of selecting an isolated polynucleotide that affects the level of expression of a CDKI polypeptide in a plant cell, the method comprising the steps of:

(a) constructing the isolated polynucleotide comprising a nucleotide sequence of at least one of 30 contiguous nucleotides derived from the isolated polynucleotide of Claim 1;

(b) introducing the isolated polynucleotide into the plant cell;

(c) measuring the level of the polypeptide in the plant cell containing the polynucleotide; and

(d) comparing the level of the polypeptide in the plant cell containing the isolated polynucleotide with the level of the polypeptide in a plant cell that does not contain the isolated polynucleotide.

12. The method of Claim 11 wherein the isolated polynucleotide consists of a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33 that codes for the polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, and 34.

13. A method of selecting an isolated polynucleotide that affects the level of expression of a CDKI polypeptide in a plant cell, the method comprising the steps of:

(a) constructing the isolated polynucleotide of Claim 1;

(b) introducing the isolated polynucleotide into the plant cell;

(c) measuring the level of the polypeptide in the plant cell containing the polynucleotide; and

(d) comparing the level of the polypeptide in the plant cell containing the isolated polynucleotide with the level of the polypeptide in a plant cell that does not contain the polynucleotide.

14. A method of obtaining a nucleic acid fragment encoding a CDKI polypeptide comprising the steps of:

(a) synthesizing an oligonucleotide primer comprising a nucleotide sequence of at least one of 30 contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33 and a complement of such nucleotide sequences; and

(b) amplifying the nucleic acid sequence using the oligonucleotide primer.

15. A method of obtaining a nucleic acid fragment encoding a CDKI polypeptide comprising the steps of:

(a) probing a cDNA or genomic library with an isolated polynucleotide comprising at least one of 30 contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33 and a complement of such nucleotide sequences;

(b) identifying a DNA clone that hybridizes with the isolated polynucleotide;

(c) isolating the identified DNA clone; and

(d) sequencing a cDNA or genomic fragment that comprises the isolated DNA clone.

16. A method for evaluating at least one compound for its ability to inhibit the activity of a cyclin dependent kinase inhibitor, the method comprising the steps of:

(a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding the cyclin dependent kinase inhibitor polypeptide, operably linked to at least one suitable regulatory sequence;

(b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results in production of the cyclin dependent kinase inhibitor encoded by the operably linked nucleic acid fragment in the transformed host cell;

(c) optionally purifying the cyclin dependent kinase inhibitor polypeptide expressed by the transformed host cell;

(d) treating the cyclin dependent kinase inhibitor polypeptide with a compound to be tested; and

(e) comparing the activity of the cyclin dependent kinase inhibitor polypeptide that has been treated with the test compound to the activity of an untreated cyclin dependent kinase inhibitor polypeptide,

thereby selecting compounds with potential for inhibitory activity.

17. A composition comprising the isolated polynucleotide of Claim 1.

18. A composition comprising the isolated polynucleotide encoding the polypeptide of Claim 10.

19. The isolated polynucleotide of Claim 1 comprising a nucleotide sequence having at least one of 30 contiguous nucleotides.

20. A method for positive selection of a transformed cell comprising:

(a) transforming a host cell with the chimeric gene of Claim 5; and

(b) growing the transformed host cell under conditions which allow expression of a polynucleotide in an amount sufficient to complement a null mutant to provide a positive selection means.

21. The method of Claim 20 wherein the host cell is a plant.

22. The method of Claim 21 wherein the plant cell is a monocot.

23. The method of Claim 21 wherein the plant cell is a dicot.

FIGURE 1

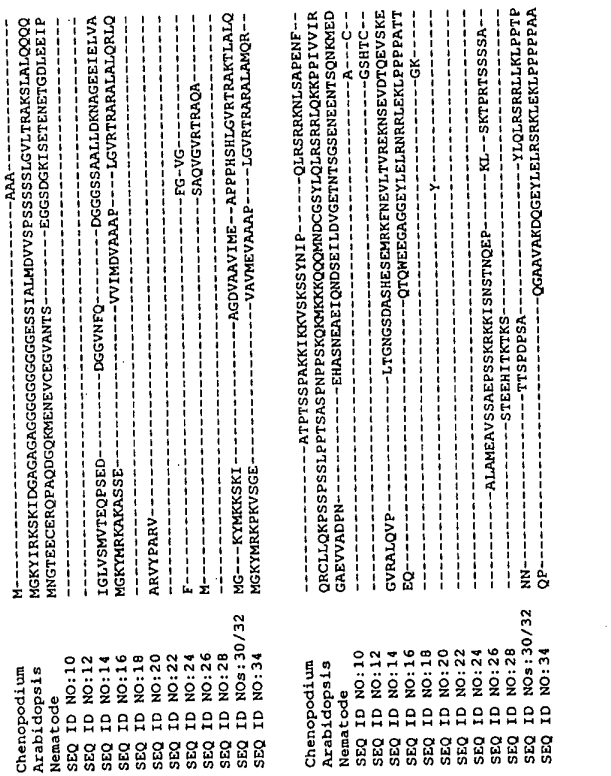

FIGURE 1

```
Chenopodium          AELETTPLEVAAVVEEEEVANCSSSEVITTARSDFPPSCCSSN----------------YD
Arabidopsis          STKRRKQQRRNETCGRNPNPRSNLDSIRGDGSRSDSVSESVVFGKDKDLISEINKDPTFG
Nematode             LEEIPVAEKVADLIEHASDEAENDNG--SDIPDMGETNTPGGGSEENTLKN---------
SEQ ID NO:10         RSDAAPAEVDGDHVPDV-VTASNSGSVPDRERRETTPSSSRAHG----------G-----
SEQ ID NO:12         RSDDA-AEAGGDHVLVVDVSASNSGSGPDRERRETTPSS-RAHG----------------
SEQ ID NO:14         MEHTDIAELQADIV----DEAAFLDC--SSIVNLVAGEETEVVNTEDGL-----------
SEQ ID NO:16         RRSGGRKAAAEAAATKEAEA-----------------------------------SYG
SEQ ID NO:18         ------------------------------------------------------------
SEQ ID NO:20         ------------------------------------------------------------
SEQ ID NO:22         ------------------------------------------------------------
SEQ ID NO:24         --------------------RC----------------------------------------
SEQ ID NO:26         VKPATVTEMVQPVSPEMVQQRC-----LSPTSSEIPASCCSSNG----------SIGLD
SEQ ID NO:28         --------------------RC-----V----------------------------------
SEQ ID NOs:30/32     PENPRRSSARNRRQISALPKP----------XXXXXXX-------XXXXXXXXXXXXX-
SEQ ID NO:34         RRRAAAAAERVEAEAEAD-EV-----------------------------------SFG

Chenopodium          QLSSSEPEVVKDDDGLGNRTADPEVESGEASSKQKESHRTEAREATKLDDQDY-PATKST
Arabidopsis          QNFFDLEEEHTQSFNRTTRESTPCSLIRRPEIMTTPG--SSTKLNICVSESNQREDSLSR
Nematode             ----KLEDLEEIPVADPIEHA-----FDGAENDNGSDIPDIGKPLHSVIVKDLSDHLVPS
SEQ ID NO:10         ----ELSDLESDLVGRQKTGCSSSP----------------------------ATTTSV
SEQ ID NO:12         ----ELSDLESDLAG-HKTG--PSL----------------------------PAATPA
SEQ ID NO:14         ----------------------------GCEEDG-----DVDKVL---------------
SEQ ID NO:16         ENMLELE-----AMERITRETTPCSLI-NTQMTSTPG--ST-RSG---HS-CHRRVNAPP
SEQ ID NO:18         -------------VGRS-------------------------------------------
SEQ ID NO:20         ---------------RNTRETTPCSLIRDPDTISTPG--STTRRS---HSSSHCKVQTPV
SEQ ID NO:22         ------------------------------------------------------------
SEQ ID NO:24         ----------------------------RREMKSSSELRENSQEPEPMEINSH-RALSKA
SEQ ID NO:26         DRIKLLDLEVESAQVETSTCNGGHEIERREMKRSSELRENSQEPEPMEINSH-RVLSKA
SEQ ID NO:28         ------------------------------------------------------------
SEQ ID NOs:30/32     ------------------RSTREGNPLLFIKGTKMPFIPPGSTTK-------PKDSPIIH
SEQ ID NO:34         ENVLESE-----AMGRGTRETTPCSLIRDSGTISTPG--STTRPS---HSNSHRRVQAPA
```


FIGURE 1

```
Chenopodium          VQIKMPSDSEIEEFFAVAEKDLQKRFSEKYNF--DIVKDVPLK--GRYDWVPI------NP
Arabidopsis          SHRRRPTTPEMDEFFSGAEEEQQKQFIEKYNF--DPVNEQPLP--GRFEWTKV------DD
Nematode             ---ILISSSDLSE--VSVQYLWSIFFWSDFVFVKPKIRNFRLKNVGIYILIHYQNRVQEKI
SEQ ID NO:10         AELIVPPAQEIQEFFAAAEAAHAKRFASKYNF--DFVRGVPLDAG-RFEWTPG---VSI--
SEQ ID NO:12         AELIVPPAHEIQEFFAAAEAAQAKRFASKYNF--DFVRGVPLDAGGRFEWAPV---VSI--
SEQ ID NO:14         ------------------------------------------------------------
SEQ ID NO:16         VHAV-PSSREMNEYFAAEQRRQQQDFIDKYNF--DPANDCPLP--GRFEWVKL------D-
SEQ ID NO:18         -----PPAEEVEAFLAAAERGMARRFAVKYNY--DIVKDAPMD-GGRYEWVRV---RP--G
SEQ ID NO:20         RHNIIPASAELEAFFAAEEQRQRQAFIDKYNF--DPVNDCPLP--GRFEWVKL------D-
SEQ ID NO:22         -----------------EEAEAKRFAAKYNF--DVVRGVPLDAG-RFEWTPV---VSSRS
SEQ ID NO:24         KA--MPTELELEEFFVAAEKDIQKRFQDKYNY--DIVKDVPLE--GRYEWVQL------KP
SEQ ID NO:26         KA--MPTELELEEFFAASEKDIQKRFQDRYNY--DIVKDVPLE--GRYEWVQL------KP
SEQ ID NO:28         -----PTESELEDFFAAAEKDIQKRFTDKYNY--DFVKDMPLE--GQYEWVKL------KS
SEQ ID NOs:30/32     EHVQRNIPTAYEMEDFFAYAEKQQQTIFMDKYNFDIVNDVPL--PGRYEWVPV---LH---
SEQ ID NO:34         RH-IIPCSAEMNEFFSAAEQPQQQAFIDKYNF--DPVNDCPLP--GRYEWVKL------D-
```

SEQUENCE LISTING

<110> E. I. du Pont de Nemours and Company

<120> Cell Cycle Genes In Plants

<130> BB1345

<140>
<141>

<150> 60/128,192
<151> 1999-04-07

<160> 39

<170> Microsoft Office 97

<210> 1
<211> 572
<212> DNA
<213> Zea mays

<220>
<221> unsure
<222> (367)

<220>
<221> unsure
<222> (474)

<220>
<221> unsure
<222> (482)

<220>
<221> unsure
<222> (537)

<220>
<221> unsure
<222> (549)

<220>
<221> unsure
<222> (560)

<220>
<221> unsure
<222> (570)

<400> 1
gcctgccgct ccgacgctgc gcctgcggag gttgacgggg atcacgtccc ggatgtcgtc 60
accgcgagca actcggggag cgtcccggac cgcgagagga gagagacgac gccatcgtcg 120
agccgggcgc acggcggcga gctcagcgat ctggagtcgg atctggtggg gcggcagaag 180
actggctgct cgtcgtcgcc ggcgacaaca acatcggttg cggagctgat cgtgccgcca 240
gcacaggaga tccaggaatt cttcgcggcc gccgaggcgg cccatgccaa acgctttgct 300
tccaagtaca acttcgactt cgtccgcggc gtgcccctcg acgccggccg gttcgagtgg 360
acgccanggg tcagcatctg aagcgagctg cggtgcaagg tgaactacta ctagctagaa 420
atgaggagaa aaaaagaaaa gatccgccaa aaaaacataa cggaaggaga aagnaaagga 480

cnttataggc aactaagcta acacatcaaa tctgctgctt ttcccggagt aactgtngaa 540
tacaggtant attacgttan cagagtactn ac 572

<210> 2
<211> 53
<212> PRT
<213> Zea mays

<400> 2
Thr Thr Ser Val Ala Glu Leu Ile Val Pro Pro Ala Gln Glu Ile Gln
1 5 10 15

Glu Phe Phe Ala Ala Ala Glu Ala Ala His Ala Lys Arg Phe Ala Ser
20 25 30

Lys Tyr Asn Phe Asp Phe Val Arg Gly Val Pro Leu Asp Ala Gly Arg
35 40 45

Phe Glu Trp Thr Pro
50

<210> 3
<211> 517
<212> DNA
<213> Oryza sativa

<400> 3
tgtatccggc acgagtgtat cggaatacca gggagacgac accttgcagc ttgatcaggg 60
accccgatac gattagcacc cctggatcta ccacaaggcg cagccactcg agttctcatt 120
gcaaggtgca aacacccgtg cgccacaaca ttattccagc atcagcagag ctggaagcgt 180
tcttcgctgc cgaagagcaa cggcaacgac aggctttcat cgacaagtat aactttgatc 240
ctgtgaatga ctgccctctt cccggccggt ttgaatgggt caagctagac tgatagattt 300
tcaggaaaag aagggcacca tggacctctc tgctccctcc acagtagtag cgtggcagag 360
gcgcttaccg tcaagttagc tttgatcctg ttgtaaaaat ttagggttag cctgtagact 420
caatggtcaa tgtgaacata cagaactgat gctgaagtta caacccctaat ccctcaacta 480
caatgtaacc cttaacaagc tcaatctgta agggaac 517

<210> 4
<211> 46
<212> PRT
<213> Oryza sativa

<400> 4
Ile Pro Ala Ser Ala Glu Leu Glu Ala Phe Phe Ala Ala Glu Glu Gln
1 5 10 15

Arg Gln Arg Gln Ala Phe Ile Asp Lys Tyr Asn Phe Asp Pro Val Asn
20 25 30

Asp Cys Pro Leu Pro Gly Arg Phe Glu Trp Val Lys Leu Asp
35 40 45

<210> 5
<211> 424
<212> DNA
<213> Glycine max

<220>
<221> unsure
<222> (274)

<220>
<221> unsure
<222> (344)

<220>
<221> unsure
<222> (350)..(351)

<220>
<221> unsure
<222> (354)

<220>
<221> unsure
<222> (361)

<220>
<221> unsure
<222> (373)

<220>
<221> unsure
<222> (379)

<220>
<221> unsure
<222> (383)

<220>
<221> unsure
<222> (391)

<220>
<221> unsure
<222> (401)

<220>
<221> unsure
<222> (405)..(406)

<220>
<221> unsure
<222> (417)

<220>
<221> unsure
<222> (419)

<220>
<221> unsure
<222> (424)

<400> 5
attcttcgtt gctgcggaga aggacattca gaaacgattt caagacaagt acaattatga 60
tattgttaag gacgtaccac tggaaggacg ctacgagtgg gttcagttga agccatgaac 120
gtgtgcgtct cgccaccgaa gaagaaaaac tccgatcaat ttgaacatgt cattttggtc 180
tatttatata ttgttaatta agtctagtct aggtctttga tttcaatctt aattatcttt 240
ttaattttac accagccaag actctttata ttcnggttgc agcttttttа aatttcggtg 300
tgattatagg taagggtaaa caggaattca gcttcgttgt tccngtcggn naancactta 360

ngctactgtg canaaaatnc ggnaattccc ngggtaagaa naaannctcc atttaangnc 420
tcan 424

<210> 6
<211> 38
<212> PRT
<213> Glycine max

<400> 6
Phe Phe Val Ala Ala Glu Lys Asp Ile Gln Lys Arg Phe Gln Asp Lys
1 5 10 15

Tyr Asn Tyr Asp Ile Val Lys Asp Val Pro Leu Glu Gly Arg Tyr Glu
20 25 30

Trp Val Gln Leu Lys Pro
35

<210> 7
<211> 609
<212> DNA
<213> Triticum aestivum

<220>
<221> unsure
<222> (540)

<220>
<221> unsure
<222> (560)

<400> 7
cgaggccgag gccgaggccg acgaggtgtc cttcggtgag aacgtgctcg agtcggaggc 60
catggggagg ggtaccaggg agacgacgcc ctgcagcttg attagggact cgggaacgat 120
aagcactcct ggatccacaa caagaccgag ccactcgaat tcccatcgca gggtgcaagc 180
tccagcgcgc catattattc catgttcagc agagatgaat gagttcttct ctgctgcgga 240
gcaaccgcaa cagcaagcct tcattgacaa gtacaacttt gatcctgtga acgactgtcc 300
tctcccaggc cgatacgagt gggtgaagct agactgataa ttctccagga aggagagcac 360
catgtacttc tccgctccct ccaccttagc gtcgtggtaa aggccgcccc gtcgtgtagc 420
tttgtttccg ttgtaaaaag aatagttagc tgtagtagcc tcaatggcgt tacatacaga 480
gtaatgctga ttacacctaa tcctcaaacc atgtacgtta caccatcgta tgacactcan 540
ggctgattta catcgtacan tcaacattat aaccactgta tgtcacacta cttatgtggt 600
gggcgatgc 609

<210> 8
<211> 46
<212> PRT
<213> Triticum aestivum

<400> 8
Ile Pro Cys Ser Ala Glu Met Asn Glu Phe Phe Ser Ala Ala Glu Gln
1 5 10 15

Pro Gln Gln Gln Ala Phe Ile Asp Lys Tyr Asn Phe Asp Pro Val Asn
20 25 30

Asp Cys Pro Leu Pro Gly Arg Tyr Glu Trp Val Lys Leu Asp
35 40 45

<210> 9
<211> 639
<212> DNA
<213> Zea mays

<400> 9
gcacgaggcc tgccgctccg acgctgcgcc tgcggaggtt gacggggatc acgtcccgga 60
tgtcgtcacc gcgagcaact cggggagcgt cccggaccgc gagaggagag agacgacgcc 120
atcgtcgagc cgggcgcacg gcggcgagct cagcgatctg gagtcggatc tggtggggcg 180
gcagaagact ggctgctcgt cgtcgccggc gacaacaaca tcggttgcgg agctgatcgt 240
gccgccagca caggagatcc aggaattctt cgcggccgcc gaggcggccc atgccaaacg 300
ctttgcttcc aagtacaact tcgacttcgt ccgcggcgtg cccctcgacg ccggccggtt 360
cgagtggacg ccaggggtca gcatctgaag cgagcgtgcg gtgcaaggtg aagctactac 420
tagctagaaa gtgaggagaa aaaaagaaaa gatgccgcca aaaaaacata acggaaagga 480
gaaaaggcaa aggagcttta taggctagct aagctaacca ccattcaaca tctgccttgc 540
ttttttctcc ggagtagact tgtagaatga cagcgtagtt atttactgtt tatccagagg 600
ttatcttgac ccccaaaaaa aaaaaaaaaa aaaaaaaaa 639

<210> 10
<211> 126
<212> PRT
<213> Zea mays

<400> 10
Ala Cys Arg Ser Asp Ala Ala Pro Ala Glu Val Asp Gly Asp His Val
1 5 10 15

Pro Asp Val Val Thr Ala Ser Asn Ser Gly Ser Val Pro Asp Arg Glu
20 25 30

Arg Arg Glu Thr Thr Pro Ser Ser Ser Arg Ala His Gly Gly Glu Leu
35 40 45

Ser Asp Leu Glu Ser Asp Leu Val Gly Arg Gln Lys Thr Gly Cys Ser
50 55 60

Ser Ser Pro Ala Thr Thr Thr Ser Val Ala Glu Leu Ile Val Pro Pro
65 70 75 80

Ala Gln Glu Ile Gln Glu Phe Phe Ala Ala Ala Glu Ala Ala His Ala
85 90 95

Lys Arg Phe Ala Ser Lys Tyr Asn Phe Asp Phe Val Arg Gly Val Pro
100 105 110

Leu Asp Ala Gly Arg Phe Glu Trp Thr Pro Gly Val Ser Ile
115 120 125

<210> 11
<211> 723
<212> DNA
<213> Zea mays

<400> 11
gcgggagcca cacctgccgc tccgacgacg ctgcagaggc tggcggggat cacgtcctgg 60
tggtggatgt ctcggcgagc aactccggga gcggcccaga ccgcgagagg agagagacga 120
cgccatcgag ccgggcgcac ggcgagctca gcgatctgga gtcggatctg gcggggcaca 180
agactggccc gtcgctaccg gcggcaacgc cggctgcgga gctgatcgtg ccgccagcac 240
acgagatcca ggagttcttc gccgccgccg aggcggccca ggccaagcgc tttgcttcca 300
agtacaactt cgacttcgtc cgcggcgtgc ccctcgacgc cggcggccgg ttcgagtggg 360

cgccggtggt cagcatctga agcgagcgtg cgtccggtgc aaggtgaagc tagaaagaga 420
aaagatgccc cccaacaaac ataacggaga agagaaaacc gaacaattaa gcagctttat 480
atagcctaag ctaaccacca ccattcatct cgtccaaatg catgccttgc ttttctctgg 540
agctagcagg agcgtagtta ttatttagta ctactttact tattcagagg ttatcttgac 600
cccgatagat caatccgctt actgtgtagt ttgtctcatg catctcttag atggagttta 660
atcgtcttaa tttaatactg tacagcagct tgctggcttg caaagaaagt ctggtttgtc 720
tct 723

<210> 12
<211> 125
<212> PRT
<213> Zea mays

<400> 12
Gly Ser His Thr Cys Arg Ser Asp Asp Ala Ala Glu Ala Gly Gly Asp
1 5 10 15

His Val Leu Val Val Asp Val Ser Ala Ser Asn Ser Gly Ser Gly Pro
20 25 30

Asp Arg Glu Arg Arg Glu Thr Thr Pro Ser Ser Arg Ala His Gly Glu
35 40 45

Leu Ser Asp Leu Glu Ser Asp Leu Ala Gly His Lys Thr Gly Pro Ser
50 55 60

Leu Pro Ala Ala Thr Pro Ala Ala Glu Leu Ile Val Pro Pro Ala His
65 70 75 80

Glu Ile Gln Glu Phe Phe Ala Ala Ala Glu Ala Ala Gln Ala Lys Arg
85 90 95

Phe Ala Ser Lys Tyr Asn Phe Asp Phe Val Arg Gly Val Pro Leu Asp
100 105 110

Ala Gly Gly Arg Phe Glu Trp Ala Pro Val Val Ser Ile
115 120 125

<210> 13
<211> 433
<212> DNA
<213> Zea mays

<220>
<221> unsure
<222> (358)

<220>
<221> unsure
<222> (415)

<400> 13
gattggtctt gtttcaatgg taaccgagca gccatctgag gacgatggag gtgtcaattt 60
tcaggatggt ggaggttcca gtgcggctct gttggacaag aatgctggtg aagagataga 120
attggtggcc ggcgttcggg ctctgcaggt accattgaca ggcaacggaa gtgatgccag 180
ccatgaaagt gaaatgagaa agttcaatga agtactaact gtcagagaga aaaacagtga 240
ggttgataca caagaagtca gtaaggagat ggaacacact gatattgctg agttgcaggc 300
tgacattgtg gatgaggcag cttttctgga ttgctcgagt attgtcaacc tggttgcngg 360
agaagagacc gaggtggtca atactgaaga tggtcttggc tgtgaggagg atggngatgt 420
tgacaaagtg ctg 433

<210> 14
<211> 144
<212> PRT
<213> Zea mays

<400> 14
Ile Gly Leu Val Ser Met Val Thr Glu Gln Pro Ser Glu Asp Asp Gly
1 5 10 15

Gly Val Asn Phe Gln Asp Gly Gly Gly Ser Ser Ala Ala Leu Leu Asp
20 25 30

Lys Asn Ala Gly Glu Glu Ile Glu Leu Val Ala Gly Val Arg Ala Leu
35 40 45

Gln Val Pro Leu Thr Gly Asn Gly Ser Asp Ala Ser His Glu Ser Glu
50 55 60

Met Arg Lys Phe Asn Glu Val Leu Thr Val Arg Glu Lys Asn Ser Glu
65 70 75 80

Val Asp Thr Gln Glu Val Ser Lys Glu Met Glu His Thr Asp Ile Ala
85 90 95

Glu Leu Gln Ala Asp Ile Val Asp Glu Ala Ala Phe Leu Asp Cys Ser
100 105 110

Ser Ile Val Asn Leu Val Ala Gly Glu Glu Thr Glu Val Val Asn Thr
115 120 125

Glu Asp Gly Leu Gly Cys Glu Glu Asp Gly Asp Val Asp Lys Val Leu
130 135 140

<210> 15
<211> 1005
<212> DNA
<213> Zea mays

<220>
<221> unsure
<222> (1001)

<400> 15
ggtccggaat tcccgggtcg acccacgcgt ccgccctcct gtgtacacca ctcccgcccc 60
gcctaccatt ttatccccgc ctctcctggc ctctgccgcc ccgtcgcaca gaatcgcttg 120
gtgcaccctg cgagggcctc ctcgaaaccc tagcttgccc agcccctccg ggccatgggc 180
aagtacatgc gcaaggccaa ggcttccagc gaggttgtca tcatggatgt cgccgccgct 240
ccgctcggag tccgcacccg agcgcgcgcc ctcgcgctgc agcgtctgca ggagcagcag 300
acgcagtggg aggaaggtgc tggcggcgag tacctggagc taaggaaccg gaggctcgag 360
aagctgccgc cgccgccggc gaccacgagg aggtcgggcg ggaggaaagc ggcagccgag 420
gccgccgcaa ctaaggaggc tgaggcgtcg tacggggaga acatgctcga gttggaggcc 480
atggagagga ttaccaggga gacgacgccc tgcagcttga ttaacaccca gatgactagc 540
actcctgggt ccacgagatc cggccactct tgccaccgca gggtgaacgc tcctccggtg 600
cacgccgtcc caagttctag ggagatgaat gagtacttcg ctgccgaaca gcgacggcaa 660
cagcaggatt tcattgacaa gtacaacttc gatcctgcaa acgactgccc tctcccaggc 720
aggtttgagt gggtgaagct agactgatgg attcagaggg acgagagagc agcaggcatg 780
gaatggaact cactccccgc tccctccaca ccacccagcg ttgtggcaga ggcgcatacc 840
gtcgtgtagc ttcgtttctg ctgtaaaaaa aaaccttagt gttttattta acatgtacct 900

taactggtct gtgtacagtc agaactgatc tgagttcaac acctgatcga tcctcactca 960
tgtacctaca ctatcgtagg acttacctgt tccgttcgac nataa 1005

<210> 16
<211> 190
<212> PRT
<213> Zea mays

<400> 16
Met Gly Lys Tyr Met Arg Lys Ala Lys Ala Ser Ser Glu Val Val Ile
1 5 10 15

Met Asp Val Ala Ala Ala Pro Leu Gly Val Arg Thr Arg Ala Arg Ala
20 25 30

Leu Ala Leu Gln Arg Leu Gln Glu Gln Gln Thr Gln Trp Glu Glu Gly
35 40 45

Ala Gly Gly Glu Tyr Leu Glu Leu Arg Asn Arg Arg Leu Glu Lys Leu
50 55 60

Pro Pro Pro Pro Ala Thr Thr Arg Arg Ser Gly Gly Arg Lys Ala Ala
65 70 75 80

Ala Glu Ala Ala Ala Thr Lys Glu Ala Glu Ala Ser Tyr Gly Glu Asn
85 90 95

Met Leu Glu Leu Glu Ala Met Glu Arg Ile Thr Arg Glu Thr Thr Pro
100 105 110

Cys Ser Leu Ile Asn Thr Gln Met Thr Ser Thr Pro Gly Ser Thr Arg
115 120 125

Ser Gly His Ser Cys His Arg Arg Val Asn Ala Pro Pro Val His Ala
130 135 140

Val Pro Ser Ser Arg Glu Met Asn Glu Tyr Phe Ala Ala Glu Gln Arg
145 150 155 160

Arg Gln Gln Gln Asp Phe Ile Asp Lys Tyr Asn Phe Asp Pro Ala Asn
165 170 175

Asp Cys Pro Leu Pro Gly Arg Phe Glu Trp Val Lys Leu Asp
180 185 190

<210> 17
<211> 472
<212> DNA
<213> Zea mays

<220>
<221> unsure
<222> (462)

<400> 17
ggaaaggtgg ggaggtcgcc gccggcggag gaggtggagg cgttcctcgc cgcagcggag 60
cgcggcatgg cgcggcgctt cgcggtcaag tacaactatg acatcgtcaa ggacgctccc 120
atggacggcg gccggtacga gtgggtccga gtgcggcccg gttaactgcc agatgcccag 180
atccaatcca agaggtcgcc tcgtgtcctc tcctgtctgt ctctacttct cttgtaaaag 240
tcgcttaagt actacccgtg taacgcttcg cttagctata actaaattat gctcacgaga 300

tgggttaatc atgtgaaggc ccaaccttgt acctacgtta gtggtgtgcg gttaacgtgt 360
ccttccttgg tccagagagg aaaagctgta acttgtctct gtagttcttc catttgtcga 420
gtctctgttc atactacatc aatgaaacag atattcaggc angcccacaa aa 472

<210> 18
<211> 54
<212> PRT
<213> Zea mays

<400> 18
Gly Lys Val Gly Arg Ser Pro Pro Ala Glu Glu Val Glu Ala Phe Leu
1 5 10 15

Ala Ala Ala Glu Arg Gly Met Ala Arg Arg Phe Ala Val Lys Tyr Asn
20 25 30

Tyr Asp Ile Val Lys Asp Ala Pro Met Asp Gly Gly Arg Tyr Glu Trp
35 40 45

Val Arg Val Arg Pro Gly
50

<210> 19
<211> 628
<212> DNA
<213> Oryza sativa

<400> 19
gcacgagtgt atccggcacg agtgtatcgg aataccaggg agacgacacc ttgcagcttg 60
atcagggacc ccgatacgat tagcacccct ggatctacca caaggcgcag ccactcgagt 120
tctcattgca aggtgcaaac acccgtgcgc cacaacatta ttccagcatc agcagagctg 180
gaagcgttct tcgctgccga agagcaacgg caacgacagg ctttcatcga caagtataac 240
tttgatcctg tgaatgactg ccctcttccc ggccggtttg aatgggtcaa gctagactga 300
tagattttca ggaaaagaag ggcaccatgg acctctctgc tccctccaca gtagtagcgt 360
ggcagaggcg cttaccgtca agttagcttt gatcctgttg taaaaattta gggttagcct 420
gtagactcaa tggtcaatgt gaacatacag aactgatgct gagttacaac cctaatccct 480
caactacaat gtaaccctta acagctcatt ctgtaaggaa ccacctcctc ctctagggcc 540
tagctagcct tatcatctgt tattaccagt tgctggatta atgaagttag atctagatat 600
tgtgtcacaa aaaaaaaaa aaaaaaaa 628

<210> 20
<211> 99
<212> PRT
<213> Oryza sativa

<400> 20
Ala Arg Val Tyr Pro Ala Arg Val Tyr Arg Asn Thr Arg Glu Thr Thr
1 5 10 15

Pro Cys Ser Leu Ile Arg Asp Pro Asp Thr Ile Ser Thr Pro Gly Ser
20 25 30

Thr Thr Arg Arg Ser His Ser Ser Ser His Cys Lys Val Gln Thr Pro
35 40 45

Val Arg His Asn Ile Ile Pro Ala Ser Ala Glu Leu Glu Ala Phe Phe
50 55 60

Ala Ala Glu Glu Gln Arg Gln Arg Gln Ala Phe Ile Asp Lys Tyr Asn
65 70 75 80

Phe Asp Pro Val Asn Asp Cys Pro Leu Pro Gly Arg Phe Glu Trp Val
85 90 95

Lys Leu Asp

<210> 21
<211> 204
<212> DNA
<213> Oryza sativa

<220>
<221> unsure
<222> (6)

<220>
<221> unsure
<222> (9)

<220>
<221> unsure
<222> (136)

<220>
<221> unsure
<222> (144)

<220>
<221> unsure
<222> (149)

<220>
<221> unsure
<222> (155)..(156)

<220>
<221> unsure
<222> (189)

<220>
<221> unsure
<222> (191)

<400> 21
cgcggnggnc gaggaggctg aggccaagcg cttcgccgcc aagtacaact tcgacgtcgt 60
tcgcggcgtg cccctcgacg ccggtcggtt cgagtggact ccggtggtca agcagccgaa 120
gctgaagcga gcgtgnagat taangcggna gctannaagg aaggtacagg gggggcgccg 180
ttgtataana nggaaaggcg agct 204

<210> 22
<211> 37
<212> PRT
<213> Oryza sativa

<400> 22
Glu Glu Ala Glu Ala Lys Arg Phe Ala Ala Lys Tyr Asn Phe Asp Val
1 5 10 15

```
Val Arg Gly Val Pro Leu Asp Ala Gly Arg Phe Glu Trp Thr Pro Val
            20                  25                  30

Val Ser Ser Arg Ser
        35
```

<210> 23
<211> 620
<212> DNA
<213> Glycine max

<400> 23

```
ttttttggtg tcggccgttg caggagagag atgaaaagtt ccagcgagct tcgagagaat  60
tcacaggagc cggagccaat ggagatcaat tctcaccgtg ccttatcaaa ggcaaaagcc 120
atgcctaccg agttggagct cgaggaattc ttcgttgctg cggagaagga cattcagaaa 180
cgatttcaag acaagtacaa ttatgatatt gttaaggacg taccactgga aggrcgctac 240
gagtgggttc agttgaagcc atgaacgtgt gcgtctcgcc accgaagaag aaaaactccg 300
atcaatttga acatgtcatt ttggtctatt tatatattgt taattaagtc tagtctaggt 360
ctttgatttc aatcttaatt atctttttaa ttttacacca gccaagactc ttttatattc 420
tggttgcagc ttttttata tttcggtgtg attataggtt aagggtaaac aggaattcag 480
cttcgtttgt tctctgtacg gagaagcagc ttaaagctag cttgtgcaga aaaatactgt 540
aaatttccct tggttaagaa gaacaaacgc ttccatttta cagacttcaa aaaaaaaaaa 600
aaaaaaaaac tcgagactag                                           620
```

<210> 24
<211> 87
<212> PRT
<213> Glycine max

<400> 24

```
Phe Phe Gly Val Gly Arg Cys Arg Arg Glu Met Lys Ser Ser Ser Glu
  1               5                  10                  15

Leu Arg Glu Asn Ser Gln Glu Pro Glu Pro Met Glu Ile Asn Ser His
            20                  25                  30

Arg Ala Leu Ser Lys Ala Lys Ala Met Pro Thr Glu Leu Glu Leu Glu
        35                  40                  45

Glu Phe Phe Val Ala Ala Glu Lys Asp Ile Gln Lys Arg Phe Gln Asp
    50                  55                  60

Lys Tyr Asn Tyr Asp Ile Val Lys Asp Val Pro Leu Glu Gly Arg Tyr
65                  70                  75                  80

Glu Trp Val Gln Leu Lys Pro
                85
```

<210> 25
<211> 1116
<212> DNA
<213> Glycine max

<400> 25

```
gcacgagaaa aaacgcacag cacagcatcg agcacagcaa cacctgaaaa aggaagggtt  60
agtgagtgtg tgtgtggaaa ttgtcatcgc tcgcaagagc aacaacaact agtacttaaa 120
gagacagaga gtgtgcacat caatgtctgc tcaggtcggt gtcaggacac gagcccaagc 180
cgcattagcc atggaagctg ttagttctgc tgaaccatca tccaagagaa agaagatcag 240
caacagtact aaccaagagc caaaactctc caagactccg agaacgagtt cttcctccgc 300
tgtcaaacca gcgacggtga cggagatggt tcagccggtg tcgccggaga tggttcagca 360
```



```
acgctgcctg agccctacct ccagtgaaat tccggcgtct tgctgctcca gcaacggatc  420
cattggcctc gatcaggaca ggatcaagct cttagatctg gaggtggaga gcgcgcaagt  480
tgaaacgtcg acgtgcaatg gtggtcatga aattgagagg agagagatga aacgttccag  540
cgagcttcgc gagaattctc aggagccgga gccaatggag atcaattctc accgtgtctt  600
atcaaaggca aaagccatgc ctaccgaatt ggagctcgag gaattcttcg ctgcctcgga  660
gaaagacatt cagaaacgat ttcaagacag atacaattat gatattgtta aggacgtacc  720
gctggaagga cgctacgagt gggttcagtt gaagccttga acgtgtgcgt cccagttcgt  780
atcgccatcg aagaagaaaa gctccgatca aattgaacat gtcattttgg cctatttata  840
tattgttaat taagtctagt ctaggtcttt gatttcaatc taaattatct ttttaattta  900
caccagccaa gactctttta tgatctggtg tacagctttt ttatatttat ttcggtgtgg  960
ttagttattg gttaagggta aacaggaatg aattcagctt cgtttgttct ctgtacggag 1020
agcagcttaa agctagcttg tgcagaaaaa tactgtaaat tccctttggt gaagaagaac 1080
aaacgcttcc attttacaaa aaaaaaaaaa aaaaaa                          1116
```

<210> 26
<211> 205
<212> PRT
<213> Glycine max

<400> 26

```
Met Ser Ala Gln Val Gly Val Arg Thr Arg Ala Gln Ala Ala Leu Ala
  1               5                  10                  15

Met Glu Ala Val Ser Ser Ala Glu Pro Ser Ser Lys Arg Lys Lys Ile
            20                  25                  30

Ser Asn Ser Thr Asn Gln Glu Pro Lys Leu Ser Lys Thr Pro Arg Thr
        35                  40                  45

Ser Ser Ser Ser Ala Val Lys Pro Ala Thr Val Thr Glu Met Val Gln
    50                  55                  60

Pro Val Ser Pro Glu Met Val Gln Gln Arg Cys Leu Ser Pro Thr Ser
65                  70                  75                  80

Ser Glu Ile Pro Ala Ser Cys Cys Ser Ser Asn Gly Ser Ile Gly Leu
                85                  90                  95

Asp Gln Asp Arg Ile Lys Leu Leu Asp Leu Glu Val Glu Ser Ala Gln
            100                 105                 110

Val Glu Thr Ser Thr Cys Asn Gly Gly His Glu Ile Glu Arg Arg Glu
        115                 120                 125

Met Lys Arg Ser Ser Glu Leu Arg Glu Asn Ser Gln Glu Pro Glu Pro
    130                 135                 140

Met Glu Ile Asn Ser His Arg Val Leu Ser Lys Ala Lys Ala Met Pro
145                 150                 155                 160

Thr Glu Leu Glu Leu Glu Glu Phe Phe Ala Ala Ser Glu Lys Asp Ile
                165                 170                 175

Gln Lys Arg Phe Gln Asp Arg Tyr Asn Tyr Asp Ile Val Lys Asp Val
            180                 185                 190

Pro Leu Glu Gly Arg Tyr Glu Trp Val Gln Leu Lys Pro
        195                 200                 205
```

<210> 27
<211> 470
<212> DNA
<213> Glycine max

<400> 27
ttcaacggag gagcatatca ccaaaaccaa atctcgctgc gttccaacgg agtcggagct 60
cgaagatttc ttcgctgctg cggagaaaga cattcagaaa cgcttcacag acaagtataa 120
ttatgatttt gtgaaggaca tgcctttgga gggacaatac gagtgggtta aattgaagtc 180
ataaaagtga aagtaccaac cttgaaagaa gaagaagaag aagaccgtta attaaaaaat 240
catgccactt gctgacatta tttttacgtt aattaattta ctttaggtac ttcatttttt 300
acttcgactt ttcgtttttt cacggtagga ctacgtttgg tgtacagtta attttatttt 360
tcttcttcta tggttaaatt ccctttgtta tttatttgtg tacggtttat gatatagtta 420
tacggagaaa gctagttctg tgaaaaaaaa aaaaaaaaaa aaaactcgag 470

<210> 28
<211> 60
<212> PRT
<213> Glycine max

<400> 28
Ser Thr Glu Glu His Ile Thr Lys Thr Lys Ser Arg Cys Val Pro Thr
1 5 10 15

Glu Ser Glu Leu Glu Asp Phe Phe Ala Ala Ala Glu Lys Asp Ile Gln
20 25 30

Lys Arg Phe Thr Asp Lys Tyr Asn Tyr Asp Phe Val Lys Asp Met Pro
35 40 45

Leu Glu Gly Gln Tyr Glu Trp Val Lys Leu Lys Ser
50 55 60

<210> 29
<211> 469
<212> DNA
<213> Glycine max

<400> 29
gcacgaggta ctgcagctgt actgtactgg ctaaaacccc aatccaagaa ccctaataaa 60
actccttccc tttcctttcc tttccattcc ccacgcctaa attctcactc cctctaaaaa 120
cccctctttc ctttcccttt ctttttctaa tttctttctt ttcgattcct gcaaaaccct 180
tacatgggca agtacatgaa gaagtccaaa atcgccggcg acgtcgccgc cgtgatcatg 240
gaggctccgc cgccgcactc ccacctcggc gtccgcaccc gcgccaagac cctcgctctc 300
cagaacaaca ccacctcccc ggacccctcc gcctacctcc agctccgcag ccggcgcctc 360
ctcaagctcc cccctacccc gccggaaaat ccccgccgct cctccgcccg aaaccgccgg 420
caaatttccg ccctccccaa acccccaaaa actttcatcc tttcaaata 469

<210> 30
<211> 87
<212> PRT
<213> Glycine max

<400> 30
Met Gly Lys Tyr Met Lys Lys Ser Lys Ile Ala Gly Asp Val Ala Ala
1 5 10 15

Val Ile Met Glu Ala Pro Pro Pro His Ser His Leu Gly Val Arg Thr
20 25 30

Arg Ala Lys Thr Leu Ala Leu Gln Asn Asn Thr Thr Ser Pro Asp Pro
35 40 45

Ser Ala Tyr Leu Gln Leu Arg Ser Arg Arg Leu Leu Lys Leu Pro Pro
50 55 60

Thr Pro Pro Glu Asn Pro Arg Arg Ser Ser Ala Arg Asn Arg Arg Gln
65 70 75 80

Ile Ser Ala Leu Pro Lys Pro
85

<210> 31
<211> 467
<212> DNA
<213> Glycine max

<400> 31
tttttgaccc cgagccccag agaagaaaag gagcaccagg gaaggcaacc ccctttttatt 60
tattaagggg actaaaatgc cattcatacc ccctggttca accaccaagc caaaggactc 120
gccaataatc catgaacacg tacaaagaaa tattccaacg gcttatgaga tggaggattt 180
ctttgcttat gctgagaagc agcaacagac aatatttatg gacaagtaca atttcgacat 240
tgtcaatgac gtacctctgc ctggacggta cgagtgggtc ccagtactcc actaggagtg 300
tcatatggtg gtgattatat atatggattg caagaaccat tcgacgtgta ttttaatttt 360
aacaactaga agaagctttc actaaccatt tagattgctt gaggctgttg tttaacaagc 420
tacaagggaa aggatctgtt tagaaatttc aaaaaaaaaa aataaaa 467

<210> 32
<211> 88
<212> PRT
<213> Glycine max

<400> 32
Arg Ser Thr Arg Glu Gly Asn Pro Leu Leu Phe Ile Lys Gly Thr Lys
1 5 10 15

Met Pro Phe Ile Pro Pro Gly Ser Thr Thr Lys Pro Lys Asp Ser Pro
20 25 30

Ile Ile His Glu His Val Gln Arg Asn Ile Pro Thr Ala Tyr Glu Met
35 40 45

Glu Asp Phe Phe Ala Tyr Ala Glu Lys Gln Gln Gln Thr Ile Phe Met
50 55 60

Asp Lys Tyr Asn Phe Asp Ile Val Asn Asp Val Pro Leu Pro Gly Arg
65 70 75 80

Tyr Glu Trp Val Pro Val Leu His
85

<210> 33
<211> 1169
<212> DNA
<213> Triticum aestivum

<400> 33
ccaggccaag cctcccctgc ctcccccatt attcccgggc ctcgtcgctc cccgctccgc 60
ctaccaattt acccgcgcct cgcttaaatc cgccaggcgc acccaggggg gcccaaaccc 120
tagcccgggc cgcgccgcgc atgggcaagt acatgcgcaa gcccaaggtc tccggcgagg 180

```
tggccgtcat ggaggtcgcc gccgcgccgc taggggtccg cacccgcgca cgagcgctcg  240
cgatgcagag gcagccgcag ggggcggcgg tggccaagga ccaggggcag tacctggagc  300
tcaggagtcg gaagctcgag aagctgcccc cgccgccgcc ggcggcgagg aggagggcgg  360
ccgcggcgga gcgtgtcgag gccgaggccg aggccgacga ggtgtccttc ggtgagaacg  420
tgctcgagtc ggaggccatg gggaggggta ccagggagac gacgccctgc agcttgatta  480
gggactcggg aacgataagc actcctggat ccacaacaag accgagccac tcgaattccc  540
atcgcagggt gcaagctcca gcgcgccata ttattccatg ttcagcagag atgaatgagt  600
tcttctctgc tgcggagcaa ccgcaacagc aagccttcat tgacaagtac aactttgatc  660
ctgtgaacga ctgtcctctc ccaggccgat acgagtgggt gaagctagac tgataattct  720
ccaggaagga gagcaccatg tacttctccg ctccctccac cttagcgtcg tggtaaaggc  780
gcgccccgtc gtgttagctt tgtttccgtt gtaaaaagaa ttaggttagc ctgtagtagc  840
ctcaatggtc gtgtaacata cagaagtaat gctgagttac accctaatcc ctcaaactcc  900
aatgtaaccg ttagcagctc attctgtaat gaccacctct agggcctgat cttatcatct  960
gttaccagtt gccagactaa tgtagtagag ctagctagat attgtatcac agcttaactt 1020
gtaatgtggg ctgggctgca gtgccatctg cattccagtg ttgtgctgca ttgcacctga 1080
agtgctgtaa cttccgtcta ttaccttctg gttcatggta tgtgaaaagg ctggtgacgg 1140
tgcttgaaaa aaaaaaaaaa aaaaaaaaa                                   1169
```

<210> 34
<211> 190
<212> PRT
<213> Triticum aestivum

<400> 34

```
Met Gly Lys Tyr Met Arg Lys Pro Lys Val Ser Gly Glu Val Ala Val
  1               5                  10                  15

Met Glu Val Ala Ala Ala Pro Leu Gly Val Arg Thr Arg Ala Arg Ala
            20                  25                  30

Leu Ala Met Gln Arg Gln Pro Gln Gly Ala Ala Val Ala Lys Asp Gln
        35                  40                  45

Gly Glu Tyr Leu Glu Leu Arg Ser Arg Lys Leu Glu Lys Leu Pro Pro
    50                  55                  60

Pro Pro Pro Ala Ala Arg Arg Arg Ala Ala Ala Ala Glu Arg Val Glu
65                  70                  75                  80

Ala Glu Ala Glu Ala Asp Glu Val Ser Phe Gly Glu Asn Val Leu Glu
                85                  90                  95

Ser Glu Ala Met Gly Arg Gly Thr Arg Glu Thr Thr Pro Cys Ser Leu
            100                 105                 110

Ile Arg Asp Ser Gly Thr Ile Ser Thr Pro Gly Ser Thr Thr Arg Pro
        115                 120                 125

Ser His Ser Asn Ser His Arg Arg Val Gln Ala Pro Ala Arg His Ile
    130                 135                 140

Ile Pro Cys Ser Ala Glu Met Asn Glu Phe Phe Ser Ala Ala Glu Gln
145                 150                 155                 160

Pro Gln Gln Gln Ala Phe Ile Asp Lys Tyr Asn Phe Asp Pro Val Asn
                165                 170                 175

Asp Cys Pro Leu Pro Gly Arg Tyr Glu Trp Val Lys Leu Asp
            180                 185                 190
```


<210> 35
<211> 196
<212> PRT
<213> Chenopodium rubrum

<400> 35

```
Met Ala Ala Ala Ala Thr Pro Thr Ser Ser Pro Ala Lys Lys Ile Lys
  1               5                  10                  15

Lys Val Ser Lys Ser Ser Tyr Asn Ile Pro Gln Leu Arg Ser Arg Arg
            20                  25                  30

Lys Asn Leu Ser Ala Pro Glu Asn Phe Ala Glu Leu Glu Thr Thr Pro
        35                  40                  45

Leu Glu Val Ala Ala Val Val Glu Glu Glu Glu Val Ala Asn Cys Ser
    50                  55                  60

Ser Ser Glu Val Ile Thr Thr Ala Arg Ser Asp Phe Pro Pro Ser Cys
65                  70                  75                  80

Cys Ser Ser Asn Tyr Asp Gln Leu Ser Ser Ser Glu Pro Glu Val Val
                85                  90                  95

Lys Asp Asp Asp Gly Leu Gly Asn Arg Thr Ala Asp Pro Glu Val Glu
            100                 105                 110

Ser Gly Glu Ala Ser Ser Lys Gln Lys Glu Ser His Arg Thr Glu Ala
        115                 120                 125

Arg Glu Ala Thr Lys Leu Asp Asp Gln Asp Tyr Pro Ala Thr Lys Ser
    130                 135                 140

Thr Val Gln Ile Lys Met Pro Ser Asp Ser Glu Ile Glu Glu Phe Phe
145                 150                 155                 160

Ala Val Ala Glu Lys Asp Leu Gln Lys Arg Phe Ser Glu Lys Tyr Asn
                165                 170                 175

Phe Asp Ile Val Lys Asp Val Pro Leu Lys Gly Arg Tyr Asp Trp Val
            180                 185                 190

Pro Ile Asn Pro
        195
```

<210> 36
<211> 256
<212> PRT
<213> Caenorhabditis elegans

<400> 36

```
Met Asn Gly Thr Glu Glu Cys Glu Arg Gln Pro Ala Gln Asp Gly Gln
  1               5                  10                  15

Lys Met Glu Asn Glu Val Cys Glu Gly Val Ala Asn Thr Ser Glu Gly
            20                  25                  30

Gly Ser Asp Gly Lys Ile Ser Glu Thr Glu Asn Glu Thr Gly Asp Leu
        35                  40                  45
```

```
Glu Glu Ile Pro Gly Ala Glu Val Val Ala Asp Pro Asn Glu His Ala
     50                  55                  60

Ser Asn Glu Ala Glu Ile Gln Asn Asp Ser Glu Ile Leu Asp Val Gly
 65                  70                  75                  80

Glu Thr Asn Thr Ser Gly Ser Glu Asn Glu Glu Asn Thr Ser Gln Asn
                 85                  90                  95

Lys Met Glu Asp Leu Glu Glu Ile Pro Val Ala Glu Lys Val Ala Asp
            100                 105                 110

Leu Ile Glu His Ala Ser Asp Glu Ala Glu Asn Asp Asn Gly Ser Asp
        115                 120                 125

Ile Pro Asp Met Gly Glu Thr Asn Thr Pro Gly Gly Gly Ser Glu Glu
    130                 135                 140

Asn Thr Leu Lys Asn Lys Leu Glu Asp Leu Glu Glu Ile Pro Val Ala
145                 150                 155                 160

Asp Pro Ile Glu His Ala Phe Asp Gly Ala Glu Asn Asp Asn Gly Ser
                165                 170                 175

Asp Ile Pro Asp Ile Gly Lys Pro Leu His Ser Val Ile Val Lys Asp
            180                 185                 190

Leu Ser Asp His Leu Val Pro Ser Ile Leu Ile Ser Ser Ser Asp Leu
        195                 200                 205

Ser Glu Val Ser Val Gln Tyr Leu Trp Ser Ile Phe Phe Trp Ser Asp
    210                 215                 220

Phe Val Phe Val Lys Pro Lys Ile Arg Asn Phe Arg Leu Lys Asn Val
225                 230                 235                 240

Gly Ile Tyr Ile Leu Ile His Tyr Gln Asn Arg Val Gln Glu Lys Ile
                245                 250                 255
```

<210> 37
<211> 327
<212> PRT
<213> Arabidopsis thaliana

```
<400> 37
Met Gly Lys Tyr Ile Arg Lys Ser Lys Ile Asp Gly Ala Gly Ala Gly
  1               5                  10                  15

Ala Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Glu Ser Ser Ile Ala
            20                  25                  30

Leu Met Asp Val Val Ser Pro Ser Ser Ser Ser Ser Leu Gly Val Leu
        35                  40                  45

Thr Arg Ala Lys Ser Leu Ala Leu Gln Gln Gln Gln Gln Arg Cys Leu
    50                  55                  60

Leu Gln Lys Pro Ser Ser Pro Ser Ser Leu Pro Pro Thr Ser Ala Ser
65                  70                  75                  80
```



```
Pro Asn Pro Pro Ser Lys Gln Lys Met Lys Lys Lys Gln Gln Gln Met
                85                  90                  95

Asn Asp Cys Gly Ser Tyr Leu Gln Leu Arg Ser Arg Arg Leu Gln Lys
            100                 105                 110

Lys Pro Pro Ile Val Val Ile Arg Ser Thr Lys Arg Arg Lys Gln Gln
        115                 120                 125

Arg Arg Asn Glu Thr Cys Gly Arg Asn Pro Asn Pro Arg Ser Asn Leu
    130                 135                 140

Asp Ser Ile Arg Gly Asp Gly Ser Arg Ser Asp Ser Val Ser Glu Ser
145                 150                 155                 160

Val Val Phe Gly Lys Asp Lys Asp Leu Ile Ser Glu Ile Asn Lys Asp
                165                 170                 175

Pro Thr Phe Gly Gln Asn Phe Phe Asp Leu Glu Glu Glu His Thr Gln
            180                 185                 190

Ser Phe Asn Arg Thr Thr Arg Glu Ser Thr Pro Cys Ser Leu Ile Arg
        195                 200                 205

Arg Pro Glu Ile Met Thr Thr Pro Gly Ser Ser Thr Lys Leu Asn Ile
    210                 215                 220

Cys Val Ser Glu Ser Asn Gln Arg Glu Asp Ser Leu Ser Arg Ser His
225                 230                 235                 240

Arg Arg Arg Pro Thr Thr Pro Glu Met Asp Glu Phe Phe Ser Gly Ala
                245                 250                 255

Glu Glu Glu Gln Gln Lys Gln Phe Ile Glu Lys Tyr Val Phe Pro Arg
            260                 265                 270

Phe Ile Cys Ser Val Leu Leu Val Met Ser Phe Gln Phe Val Leu Phe
        275                 280                 285

Phe Ser Phe Gly Leu Val Ser Leu Met Val Ser Val Asn Ser Phe Phe
    290                 295                 300

Arg Tyr Asn Phe Asp Pro Val Asn Glu Gln Pro Leu Pro Gly Arg Phe
305                 310                 315                 320

Glu Trp Thr Lys Val Asp Asp
                325
```

<210> 38
<211> 870
<212> DNA
<213> Arabidopsis thaliana

```
<400> 38
atggggaaat acataagaaa gagcaaaatc gacggagctg gagctggagc tggtggtgga  60
ggtggaggcg gaggtggagg tgaatcctct attgctttaa tggatgttgt ttcacctttct 120
tcctcttctt cacttggtgt tttaactcga gctaaatctc tagctcttca acaacaacaa 180
caacgctgtc ttcttcaaaa accctcttct ccttcttcgt taccaccgac ttctgcttct 240
cctaatccac cgtcaaagca gaagatgaag aagaagcagc agcagatgaa cgattgtggt 300
```

```
tcgtatcttc agctacggag tcgtcggctt cagaagaaac caccaattgt tgtgattcgt 360
tctactaaac ggaggaaaca gcaacggagg aatgagacat gcggaagaaa ccctaaccct 420
aggagtaatt tggattcgat tcgtggtgat gggtctaggt ctgattctgt ttctgaaagt 480
gttgtctttg gtaaagacaa ggatttgatt tctgaaatca ataaagatcc aacctttggc 540
caaaatttct tcgaccttga agaagaacac actcaaagct tcaacaggac cacaagggaa 600
tccacaccat gcagtttgat aagaaggcct gaaatcatga caaccccggg atcgtctaca 660
aagcttaaca tttgtgtaag tgaaagcaac cagagagaag acagtttgtc acgcagccat 720
cgtcgccgac caactacacc tgaaatggac gagtttttct cgggtgctga agaagaacaa 780
cagaagcaat tcattgagaa gtacaacttt gatcctgtga acgaacaacc actaccagga 840
cggtttgaat ggacgaaggt agatgattag                                  870
```

<210> 39
<211> 289
<212> PRT
<213> Arabidopsis thaliana

<400> 39

```
Met Gly Lys Tyr Ile Arg Lys Ser Lys Ile Asp Gly Ala Gly Ala Gly
  1               5                  10                  15

Ala Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Glu Ser Ser Ile Ala
            20                  25                  30

Leu Met Asp Val Val Ser Pro Ser Ser Ser Ser Ser Leu Gly Val Leu
        35                  40                  45

Thr Arg Ala Lys Ser Leu Ala Leu Gln Gln Gln Gln Gln Arg Cys Leu
    50                  55                  60

Leu Gln Lys Pro Ser Ser Pro Ser Ser Leu Pro Pro Thr Ser Ala Ser
65                  70                  75                  80

Pro Asn Pro Pro Ser Lys Gln Lys Met Lys Lys Lys Gln Gln Gln Met
                85                  90                  95

Asn Asp Cys Gly Ser Tyr Leu Gln Leu Arg Ser Arg Arg Leu Gln Lys
            100                 105                 110

Lys Pro Pro Ile Val Val Ile Arg Ser Thr Lys Arg Arg Lys Gln Gln
        115                 120                 125

Arg Arg Asn Glu Thr Cys Gly Arg Asn Pro Asn Pro Arg Ser Asn Leu
    130                 135                 140

Asp Ser Ile Arg Gly Asp Gly Ser Arg Ser Asp Ser Val Ser Glu Ser
145                 150                 155                 160

Val Val Phe Gly Lys Asp Lys Asp Leu Ile Ser Glu Ile Asn Lys Asp
                165                 170                 175

Pro Thr Phe Gly Gln Asn Phe Phe Asp Leu Glu Glu Glu His Thr Gln
            180                 185                 190

Ser Phe Asn Arg Thr Thr Arg Glu Ser Thr Pro Cys Ser Leu Ile Arg
        195                 200                 205

Arg Pro Glu Ile Met Thr Thr Pro Gly Ser Ser Thr Lys Leu Asn Ile
    210                 215                 220
```



```
Cys Val Ser Glu Ser Asn Gln Arg Glu Asp Ser Leu Ser Arg Ser His
225                 230                 235                 240

Arg Arg Arg Pro Thr Thr Pro Glu Met Asp Glu Phe Phe Ser Gly Ala
            245                 250                 255

Glu Glu Glu Gln Gln Lys Gln Phe Ile Glu Lys Tyr Asn Phe Asp Pro
        260                 265                 270

Val Asn Glu Gln Pro Leu Pro Gly Arg Phe Glu Trp Thr Lys Val Asp
    275                 280                 285

Asp
```